

Nonlinear Dynamics of the Voice: Bifurcations and Mode Analysis of Complex Spatio-Temporal Signals

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
Humboldt-Universität zu Berlin

von
Herr Dipl.-Phys. Jürgen Neubauer
geboren am 18.10.1972 in Pegnitz

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jürgen Mlynek

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Dr. Michael Linscheid

Gutachter:

1. Prof. David A. Berry, PhD
2. Prof. Dr. Harald Engel
3. Prof. Dr. Hanspeter Herzel

eingereicht am:	9. Februar 2004
Tag der mündlichen Prüfung:	12. August 2004

Abstract

In this thesis, the physics of phonation was discussed using the theory of nonlinear dynamics. Digital high speed recordings of human and nonhuman laryneal oscillations, image processing, signal analysis, and modal analysis have been used to quantitatively describe nonlinear phenomena in pathological human phonation, healthy voices in singing, and nonhuman mammalian larynges with vocal membranes. Bifurcation analysis of a simple mathematical model for vocal folds with vocal membranes allowed a qualitative 'nonlinear fit' of observed vocalization patterns in nonhuman mammals.

The main focus of the present work was on:

- classification of vocalizations of contemporary vocal music to provide insight to production mechanisms of complex sonorities in artistic contexts, especially to nonlinear source-tract coupling
- pathological voice instabilities induced by asymmetries within single vocal folds and between vocal folds
- dynamic effects of thin, lightweight, and vibrating vocal membranes as upward extensions of vocal folds in nonhuman mammals

In Chapter 3, complex and multiphonic voice signals of vocal improvisors associated with contemporary music were analyzed within the framework of nonlinear dynamics. Narrow-band spectrograms of complex vocalizations were used in terms of spectral bifurcation diagrams to classify nonlinear phenomena. It was shown that nonlinear phenomena such as subharmonic oscillations, toroidal oscillations, and abrupt onset of irregular chaotic oscillation are extensively used by singers in contemporary vocal music. Biphonation and triphonation (voice signal with two or three independent pitch melodies, respectively) could be found during glottal whistle production. Examples of vocalizations were shown where slowly varying formant frequencies induced abrupt transitions to subharmonics and biphonation in the vocal output when they matched harmonics of the voice source signal. This showed that nonlinear source-tract coupling is one possible mechanism for the control of voice instabilities in musical contexts. Additionally, from previous research on pathological voice production it is known that desynchronization of vocal fold oscillations could be induced by asymmetries in the larynx. Showing the recurrent use of nonlinear phenomena in consecutive voice samples of performers, it could be concluded that singers with normal healthy larynges are able to induce voice instabilities in a reproducible way for musical tasks.

Desynchronization of vibratory modes of the vocal folds was studied in Chapter 4. In patients with vocal fold pathologies, the superior edges of the oscillating vocal folds were observed *in vivo* with a high speed camera. The irregular spatio-temporal vibration patterns were described quantitatively with multi-line kymograms, spectral analysis, and spatio-temporal plots. Empirical orthogonal eigenfunctions – empirical modes of the vibrating vocal folds – were extracted from the complex spatio-temporal vibration patterns. With this decomposition technique, decorrelated coherent structures (vibratory modes) could be identified. The major finding of this analysis was that laryngeal biphonation (laryngeal oscillation with two independent pitches) could be induced either by left-right (LR) asymmetry of the two vocal folds or by desynchronized anterior-posterior (AP) vibratory modes. For the latter case, the term 'AP biphonation' was introduced to differentiate from 'LR biphonation' known from literature. As a reference, spatio-temporal data from one normal healthy voice was used for comparison. The analyzed phonation examples showed that for normal phonation the first two modes were sufficient to explain the overall glottal dynamics. The spatio-temporal oscillation patterns associated with LR biphonation could be explained by the first three modes. For AP biphonation, higher order modes were necessary to describe the oscillatory patterns. For clinical research, an entropy measure was introduced to quantify spatial irregularity. It was found that spatial irregularity was significantly higher for the voices with pathological vocal fold oscillations than for normal healthy phonation. For future modeling work, two asymmetry measures were introduced: Left-right

asymmetry was defined by the ratio of the fundamental frequencies of left and right vocal fold oscillations. Anterior-posterior asymmetry was defined by the ratio of the fundamental frequencies of first and second vibratory modes within one vocal fold. These quantities clearly differentiated between LR biphonation and AP biphonation. The findings of desynchronized oscillatory modes within the vocal fold oscillators agreed with previous theoretical work on computer models of vocal folds and *in vitro* experiments.

In Chapter 5, a phenomenological model for sound production in nonhuman mammals was studied. The well-known simplified two-mass model was extended by oscillating vocal membrane plates attached to the upper edge of the vocal folds. In nonhuman mammals, vocal membranes are one widespread morphological variation of vocal folds. In bats they are responsible to produce ultrasonic echolocation calls. In nonhuman primates they facilitate the production of highly diverse vocalizations. It was shown that in bats and primates, complex vocalizations are a crucial feature of the call repertoire. Vocalizations with subharmonic oscillations, biphonation, irregular chaotic calls, register jumps, pulsed high frequency oscillations, and abrupt transitions between these different behaviors can be found frequently. A vocal membrane model was developed to understand the production of these complex calls. One modeling goal was to keep the number of new parameters as small as possible. Thus, the vocal membrane geometry was specified by the vocal membrane height, its mass, the resting angle (overhang into the glottis) and parameters for the stiffness and viscous tissue damping. As in the two-mass model, the air flow through the model was described with the Bernoulli equation. This potential law for the driving air flow was combined with the assumption of jet separation at the most narrow point in the model glottis. Linear eigenmode analysis was used to study the influence of the vocal membrane geometry on eigenfrequencies and eigenmodes. On the background of the theory of coupled oscillators, ratios of eigenfrequencies can be interpreted in analogy to the frequency ratio in the generic circle map. Parameter regions with frequency locking (mode entrainment) could be identified in the vocal membrane model. The shape of the eigenmodes was an indicator for the phonation onset behavior of the full nonlinear model. It was found that the vocal membrane height had the strongest influence on the eigenmode shape. The eigenfrequencies were mostly effected by the frequency ratio of the model mass oscillators associated with the vocal folds and the vocal membrane oscillator. With Hopf bifurcation analysis, the influence of the vocal membrane geometry on phonation onset pressure and onset frequency was studied. The most important finding was that two voice registers of the vocal membrane model existed. The register areas could overlap when the vocal membrane height was small enough. Overlapping register areas were interpreted as indicating biphonation (toroidal oscillations) in the model. Numerical integration

with slowly varying model parameters supported this finding. Hopf bifurcation analysis also showed that the vocal membrane geometry could minimize phonation onset pressure and enlarge the phonatory pressure range of the model. As a final finding, numerical simulations with gliding subglottal pressure variations and varying frequency tuning (related to the ratio of the oscillation frequencies of vocal folds and the vocal membrane) revealed instabilities that qualitatively resembled observed vocalization patterns in bats and primates.

To date, detailed data on larynges of nonhuman primates and bats is missing. Therefore, this work is only a first step towards a quantitative 'nonlinear fit' of nonhuman vocalizations. A detailed fit of specific vocalizations is still to be done. In the present work, parameters had to be guessed. Due to the nonlinear nature of the vocal membrane model, there is no unique mapping of observed attractor states to the parameter space of a vocal membrane model. In future, the combination of modeling and experiments should continue to improve parameter estimation for the vocal membrane model and gain deeper understanding of the diversity of the nonhuman mammalian vocalization repertoire.

Keywords:

synchronization, biphonation, contemporary vocal music, echolocation

Zusammenfassung

Die Physik der Lauterzeugung (Phonation) wurde mit Hilfe der Theorie der Nichtlinearen Dynamik untersucht. Digitale Hochgeschwindigkeitsaufnahmen von Schwingungen in menschlichen und nichtmenschlichen Kehlköpfen, digitale Bildanalyse, Signalanalyse und Modenanalyse wurden zur quantitativen Beschreibung nichtlinearer Phänomene eingesetzt. Es wurden nichtlineare Phänomene bei stimmkranker (pathologischer) menschlicher Lauterzeugung untersucht, wie auch in stimmgesunden Singstimmen und in Kehlköpfen von nichtmenschlichen Säugetieren mit Stimmlippen-Membranen. Durch Bifurkationsanalyse eines einfachen mathematischen Modells für Stimmlippen mit Membranen konnten beobachtete Lautmuster nichtmenschlicher Säugetiere qualitativ “nichtlinear gefittet” werden.

Die Schwerpunkte der vorliegenden Arbeit waren:

- die Klassifikation von Lautmustern in zeitgenössischer Vokalmusik, um Erzeugungsmechanismen für komplexe Stimmklänge zu erklären, die im künstlerischen Kontext vorkommen; im besonderen war die Rolle der Quelle-Trakt-Kopplung von Interesse
- Instabilitäten in Stimmpatienten, die durch Asymmetrien in einzelnen Stimmlippen wie auch zwischen den Stimmlippen verursacht wurden.
- dynamische Effekte von dünnen, leichten und schwingenden Stimmlippen-Membranen, vertikalen Fortsätzen der Stimmlippen in nichtmenschlichen Säugetieren.

In Kapitel 3 wurden komplexe und polyphone Stimmsignale von Stimmimprovisatoren – “Vocal Improvisors”, Sängern zeitgenössischer Musik – mit Methoden der Nichtlinearen Dynamik untersucht. Schmalband-Spektrogramme komplexer Laute wurden als spektrale Bifurkationsdiagramme zur Klassifikation nichtlinearer Phänomene verwendet. Es zeigte sich, dass nichtlineare Phänomene, wie subharmonische Oszillationen, toroidale Schwingungen und abrupte Einsätze von irregulärer chaotischer Oszillation, intensiv von Sängern zeitgenössischer Vokalmusik eingesetzt werden. Die Wechselwirkung von Vokaltraktresonanzen mit der glottalen Stimmquelle konnte als ein möglicher Mechanismus für kontrollierte Stimminstabilitäten in Sängern identifiziert werden. Zusammen mit Kehlkopfasymmetrien, die Stimmlippenschwingungen desynchronisieren können, erlauben es diese Mechanismen stimmgesunden Sängern, Stimminstabilitäten gezielt und reproduzierbar für künstlerische Zwecke einzusetzen.

Die Desynchronisation von Schwingungsmoden der Stimmlippen wurde in Kapitel 4 untersucht. Stimmlippenschwingungen von Stimpatienten wurden mit einer Hochgeschwindigkeitskamera aufgenommen. Irreguläre raum-zeitliche Schwingungsmuster der Stimmlippenkanten konnten mit Empirischen Orthogonalen Eigenfunktionen (EOF) in dekorrelierte Schwingungsmoden zerlegt werden. Dadurch konnte ein neuer Mechanismus für laryngeale Biphonation (Lauterzeugung mit zwei unabhängigen Tonhöhen) *in vivo* identifiziert werden: desynchronisierte Schwingungsmoden entlang der anterior-posterior Richtung einer einzelnen Stimmlippe. Die EOF Analyse ermöglichte die Abgrenzung von der seit längerem bekannten Desynchronisierung der rechten und linken Stimmlippen aufgrund von Kehlkopfasymmetrien (LR-Biphonation). Bei dieser neuen Art der Biphonation, hier AP-Biphonation genannt, wurden deutlich mehr Schwingungsmoden der Stimmlippen angeregt als bei sowohl normaler Phonation als auch LR-Biphonation. Für den Einsatz in der klinische Diagnose und für zukünftige Stimmlippenmodelle wurden sowohl die räumlichen Irregularität der Stimmlippenschwingungen, die Rechts-Links-Asymmetrie wie auch die Anterior-Posterior-Asymmetrie quantitativ bestimmt. Diese Messparameter erlaubten die eindeutige Trennung von Stimpathologien mit LR Biphonation von solchen mit AP Biphonation.

In Kapitel 5 wurde ein phänomenologisches Modell zur Untersuchung der Lauterzeugung bei nichtmenschlichen Säugetieren, im speziellen bei Primaten und Fledermäusen, entwickelt. Dazu wurde ein vereinfachtes Zweimassenmodell um oszillierende Stimm-Membrane erweitert. Diese Stimm-Membrane finden sich in Kehlköpfen von Fledermäusen und Primaten, wo sie einerseits zur Ultraschallerzeugung verwendet werden und andererseits für eine grosse Lautvielfalt sorgen. Nicht-lineare Phänomene, wie subharmonische Oszillationen, Biphonation, irreguläres chaotisches Verhalten, Registersprünge, gepulste hochfrequente Oszillationen und abruptes Übergänge zwischen diesen Phänomenen sind gehäuft im Lautrepertoire

von Primaten und Fledermäusen zu finden. Das Stimm-Membrane-Modell wurde entwickelt, um dieses diverse Lautrepertoire mit möglichst wenigen Modellparametern zu reproduzieren. Mit linearer Eigenmodenanalyse wurden Parameterbereiche gefunden, wo nahe benachbarte Eigenfrequenzen Frequenzlocking (Modensynchronisation) des vollen nichtlinearen Modells vermuten liessen. Dieses Verhalten ist aus der Theorie gekoppelter Oszillatoren und der generischen Kreisabbildung bekannt. Die Eigenmoden des Modells liessen Rückschlüsse auf das Stimmeinsatzverhalten des vollen nichtlinearen Modells zu. Mittels Hopf-Bifurkationsanalyse wurden die Auswirkungen der Stimm-Membran auf den Stimmeinsatz (in Bezug auf subglottalem Einsatzdruck und Frequenz am Einsattpunkt) bestimmt. Das wichtigste Resultat dieser Analyse war das Vorhandensein zweier Stimmregister. Die Parameterregionen dieser Register konnten sich überschneiden, falls die Stimm-Membrane kurz genug waren. Dieses Verhalten gab Hinweise auf die Existenz von Biphonation (toroidale Oszillationen) im Modell. Numerische Simulationen bestätigten diese Vermutung. Hopf-Bifurkationsanalyse zeigte auch, dass die Geometry der Stimm-Membrane so gewählt werden konnte, dass der subglottale Einsatzdruck minimal wurde und der Druckbereich für Phonationen vergrössert wurde. Numerische Simulationen demonstrierten, dass das phänomenologische Stimm-Membran-Modell das Lautrepertoire von Fledermäusen und Primaten qualitativ reproduzieren konnte. Eine quantitative und detaillierte Anpassung des Modells steht noch aus, da die dafür benötigten Daten (Hochgeschwindigkeitsaufnahmen, Gewebisdämpfung, glottale Ruhefläche, etc.) zur Zeit noch nicht vorhanden sind.

Schlagwörter:

Synchronisation, Biphonation, Zeitgenössische Vokalmusik, Echolokation

Contents

Contents	ix
List of Figures	xii
List of Tables	xxiv
1 Introduction	1
2 Physics of Voice Production	7
2.1 Anatomy and physiology of the voice production apparatus	7
2.2 Mechanism of sound production	13
2.3 Nonlinear dynamics of phonation	17
2.3.1 Coupled oscillators – the core of phonatory dynamics	25
2.4 Experimental data acquisition of phonatory system	27
2.5 Modeling voice production in humans and nonhuman mammals	32
2.5.1 The two-mass model as basic model	35
3 Nonlinear Phenomena in Contemporary Vocal Music	41
3.1 Introduction	42
3.1.1 Nonlinear phenomena	43
3.2 Musical phenomena	48
3.3 Material and Methods	49
3.4 Results	50
3.5 Discussion	55
3.5.1 Bifurcation analysis	56
3.5.2 Physiological mechanisms	56
3.5.3 Reproducibility, intention, and control	58
3.5.4 Musical relevance	58
4 Spatio-temporal Analysis of Irregular Vocal Fold Oscillations	59
4.1 Introduction	60
4.2 Material and Methods	62

CONTENTS

4.2.1	Multi-line kymography	63
4.2.2	Time series preprocessing	65
4.2.3	Calculation of the empirical orthogonal functions	72
4.3	Analysis of spatio-temporal glottal patterns	73
4.3.1	Spectral analysis of spatio-temporal plots	73
4.3.2	EOF analysis of spatio-temporal vibration patterns	78
4.4	Discussion	84
4.5	Conclusion	92
5	Vocal Membrane Model for Bats and Primates	94
5.1	Introduction	95
5.1.1	Significance of a biomechanical vocal membrane model	95
5.1.2	Vocal membranes in nonhuman mammalian larynges	96
5.1.3	Vocal membranes in bats	96
5.1.4	Vocal membranes in squirrel monkeys	99
5.1.5	<i>In vivo</i> video imaging of oscillating vocal membranes	103
5.1.6	Biomechanical model with rigid reed-like vocal membranes	107
5.1.7	Dynamical vocal membrane model	107
5.2	Materials and Methods	108
5.2.1	Phenomenological model of vocal membrane systems	108
5.2.2	Description of vocal membrane model	111
5.2.3	Vocal membrane model equations	112
5.2.4	Biologically relevant control parameters	119
5.2.5	Numerical analysis of the vocal membrane model	121
5.3	Results	124
5.3.1	Linear eigenmode analysis	124
5.3.2	Hopf bifurcation and voice onset	132
5.3.3	Bifurcation diagrams and registers	135
5.3.4	Simulation of complex vocalizations	140
5.4	Discussion	149
5.4.1	Registers in larynges with vocal membranes	150
5.4.2	Diversity of vocal membrane dynamics	151
5.4.3	Adaptability and robustness of vocal membrane systems	152
5.5	Outlook	153
	Bibliography	155
	A Empirical Orthogonal Function Analysis	179

CONTENTS

B	Experimental Observations on Vocal Membranes	182
B.1	Selected experimental studies in bats	182
B.1.1	Subglottic pressure and the control of phonation by the echolocating bat, <i>Eptesicus</i>	184
B.1.2	Mechanisms of sound production in echolocating bats	185
B.1.3	The production of echolocation signals by bats and birds	186
B.2	Selected experimental studies in primates	187
B.2.1	Effects of laryngeal nerve transection on squirrel monkey calls	188
B.2.2	Modes of vocal variation in Sykes's monkey (<i>Cercopithecus albogularis</i>) squeals	189
B.2.3	Voicing biomechanics and squirrel monkey (<i>Saimiri boliviensis</i>) vocal communication	190
C	Derivation of vocal membrane model equations	193
C.1	Derivation of equations from first principles	193
C.2	Derivation of aerodynamic driving forces and torques	198

List of Figures

2.1	Components in the airway system in the head, neck, and chest (Titze, 1994a): The lungs produce pressure that drives the subglottal airstream. The airstream is fed to the larynx via bronchi and trachea. The primary function of the larynx is to protect the airway system from foreign material (such as food) which passes into the esophagus during swallowing. The vocal tract, comprised of pharynx and oral and nasal cavities, filters the primary sound signal generated by airstream-driven oscillations of the vocal folds in the larynx.	8
2.2	Sideview of the skeletal laryngeal framework (Titze, 1994a): Two tracheal rings, the cricoid and the thyroid cartilages, the hyoid bone, and the epiglottis are shown. Ligaments and intrinsic muscles connect the cricoid and thyroid cartilages. Cricothyroid muscle contraction rotates the cricoid cartilage against the thyroid cartilage, which lengthens the vocal folds. The cricoarytenoid muscle contraction rocks (rotates and translates) the arytenoid cartilages to open and close the glottis.	9
2.3	Superior view of the muscular and cartilagenous framework of the larynx (transverse section at the level of the vocal folds) (Titze, 1994a): The vocalis muscle (thyroarytenoid muscle) is inserted between the arytenoid and thyroid cartilages. The arytenoid cartilages can rock and move on top of the cricoid cartilage due to contraction of the thyroarytenoid muscle, the cricoarytenoid muscles and the interarytenoid muscle. The top side of the cricoid cartilage supporting the arytenoid cartilages can be moved back and forth by cricothyroid muscle and thyroarytenoid muscle contraction, respectively.	10

LIST OF FIGURES

2.4	Schematic of a frontal section through one vocal fold (Titze, 1994a): The tissue layers can be roughly grouped into vocal fold body and vocal fold cover. The body consists of the muscle fibers of the thyroarytenoid muscle. The cover is built by the epithelium and the lamina propria (superficial, intermediate, and deep layer). . . .	11
2.5	Innervation of the larynx (Titze, 1994a): Two branches of the vagus nerve innervate the intrinsic laryngeal muscles: The superior laryngeal nerve only innervates the cricothyroid muscle. The recurrent laryngeal nerve innervates all other intrinsic muscles (including agonistic-antagonistic muscle pairs).	12
2.6	Frontal section of the larynx through the center of the glottis: The vocal folds are shown in prephonatory rest position. Adduction muscle activity narrows the glottis (the space between vocal folds), thus increasing subglottal pressure. Aerodynamical forces blow these visco-elastic orifice apart, until elastic tissue recoil and aerodynamical Bernoulli forces close the glottis closes. [The ventricular folds, a valve structure similar to the vocal folds, are shown above the vocal folds. Note that in certain phonatory modes they have been observed to oscillate with the vocal folds (Fuks et al., 1998). Their aerodynamical influence of facilitating phonation in whispered voices has been shown by Miller et al. (1988). de Oliveira Rosa et al. (2003) demonstrated that they could increase the transglottal pressure drop.]	14
2.7	Schematic sideview of one glottal cycle: After prephonatory closing of the glottis, rising subglottal pressure opens vocal folds starting from below. The upper edge follows the motion of the lower edge with a certain phase shift. Due to increasing elastic tissue recoil first the lower portion of the vocal folds starts closing again. The upper portion again follows with a certain delay. (after Schönhärl, 1960)	16

LIST OF FIGURES

- 2.8 Spectrogram of sustained phonation of a trained voice with various subharmonic components: In the above time-frequency plot, the subsequent power spectra calculated from moving time segments are shown. The normalized sound intensity distributions (on a logarithmic scale) are transformed to distributions of gray scale values. Here, black points indicate the highest intensities. This spectrogram shows a segment of sustained phonation after voice onset, induced by a Hopf bifurcation from the prephonatory fixed state. Period doubling bifurcations inducing subharmonic oscillations at small integer ratios of the original fundamental frequency are visualized (see arrows for “Subharmonic components”). As in the above spectrogram, subharmonic oscillations show up as additional parallel bands of high intensity within the stack of harmonic frequencies, integer multiples of the fundamental frequency (here at about 500 Hz). The gray shading between the harmonics and subharmonics is due to noise components. 21
- 2.9 Spectrogram of sustained phonation revealing simultaneous unrelated pitch melodies $f(t)$ and $g(t)$ in the voice of a single singer. Toroidal oscillations of the singer’s phonatory system are shown: Two frequency bands with high intensities (here at about 800 Hz) vary uncorrelated over time. Short segments of entrained oscillation, associated with folded limit cycle oscillation, are interspersed: During short time segments the uncorrelated behavior of the pitch melodies $f(t)$ and $g(t)$ abruptly changes to entrained oscillations (e.g., at 0.5 sec). Frequency components at linear combinations of f and g appear in the spectrogram due to nonlinear coupling of the two independent oscillations. 22
- 2.10 Spectrogram of sustained phonation of the voice of a singer showing an abrupt onset of chaotic irregular oscillations: Following voice onset, a frequency jump at about 1.1 sec occurs during the increase of the fundamental frequency. The harmonic oscillation behavior changes to irregular noise-like behavior. Embedded in the broadband chaotic segment, a remnant of the previous fundamental frequency is visible. This is due to the presence of the previously stable limit cycle oscillation. The above example shows harmonic windows with folded limit cycle oscillations interspersed in the chaotic segment. 23

LIST OF FIGURES

2.11	Spectrogram of sustained phonation of the voice of a singer demonstrating the repetitive intentional use of nonlinear phenomena: The left part (0.0 – 2.0 sec) shows transistions from limit cycle to chaos, followed by subharmonic oscillations and again chaos. The right part (2.0 – 4.0 sec) is qualitatively similar.	24
2.12	Motion on a torus in phase space. For rational frequency ratios $\Omega = \omega_1/\omega_2$ the trajectories close after a finite number of cycles (mode-locked state). For irrational frequency ratio such motion is called quasiperiodic. The trajectory never closes and asymptotically covers the whole torus, never repeating a point (from Schuster (1988)).	26
2.13	Schematical bifurcation diagram of the circle map shows mode locking within the Arnold tongues (hatched areas) for small coupling $K < 1$, merging of Arnold tongues for critical coupling $K = 1$. Note that chaos and mode locked states coexist for coupling $K > 1$. The lines (in the region $K > 1$) indicate mode locked states associated with mode locking regions (from Schuster (1988)).	26
2.14	Sketch of the simplified two-mass model after Steinecke and Herzel (1995): lower and upper masses m_1 and m_2 , viscous damping r_1 and r_2 , elastic recoil k_1 and k_2 , coupling stiffness k_c , thickness of upper and lower masses d_1 and d_2 , rest position of lower and upper masses x_{01} and x_{02} , and displacement of masses from their resting position $x_1(t)$ and $x_2(t)$	36
3.1	Spectral bifurcation diagram (spectrogram, sonagram) revealing a period doubling cascade from a harmonic behavior with a pitch of about 500 Hz to subharmonics at $\frac{1}{2}$, $\frac{1}{4}$ or $\frac{1}{6}$ of the pitch. The sample was taken from a recording of the solo vocal improvisation <i>investigazioni (diplofonie e triplofonie)</i> by the vocal improvisor Demetrio Stratos (Stratos, 1978).	45
3.2	Spectral bifurcation diagram showing biphonation with two independent frequencies (two independent pitched melodies) termed f and g and various linear combinations of these frequencies. The sample is from a recording of the solo vocal improvisation <i>passaggi 1,2</i> by the vocal improvisor Demetrio Stratos (Stratos, 1978). . . .	46
3.3	Spectral bifurcation diagram displaying the transition from harmonic behavior via period doubling, period quadrupling to an irregular, noise-like segment. Within the broadband noise-like segment the vocal tract filter is revealed by different shadings. The sample was taken from a recording of the solo vocal improvisation <i>entre nosotros - epitafio a las ballenas</i> by the vocal improvisor Fatima Miranda (Miranda, 1992).	47

LIST OF FIGURES

3.4	Spectral bifurcation diagram showing coincidence of formant matching harmonic associated with sideband modulations of harmonic component. Here, the dynamical range for the gray scale coding of power spectrum intensities was chosen as 90 dB sound intensity level. The sample was taken from a recording of the solo vocal improvisation <i>in principio</i> by the vocal improvisor Fatima Miranda (Miranda, 1992).	51
3.5	Spectral bifurcation diagram revealing recurrent source-formant interaction: Formant-induced P5 bifurcation with five subharmonics vanishes and reappears when a formant frequency matches the fourth or the third harmonic, respectively. The dynamical range for the gray scale coding of power spectrum intensities was chosen as 90 dB sound intensity level. The figure contrast and intensity was manually adjusted to emphasize formants. The sample was taken from a recording of the solo vocal improvisation <i>investigazioni (diplofonie e triplofonie)</i> by the vocal improvisor Demetrio Stratos (Stratos, 1978).	52
3.6	Spectral bifurcation diagrams showing recurrent instances of abrupt transitions from regular phonation to subharmonics and irregular phonation (arrows) within the same vocalization sequence. The samples were taken from a recording of the solo vocal improvisation <i>entre nosotros - epitafio a las ballenas</i> by the vocal improvisor Fatima Miranda (Miranda, 1992).	53
3.7	Spectral bifurcation diagram displaying a female glottal whistle with biphonation and triphonation. In the final segment triphonation is observed with two glottal whistle components and a third, vocal fry-like oscillator. The sample was taken from a recording of the solo vocal improvisation <i>Signals</i> by the vocal improvisor Anna Homler (Homler, 1989).	54
4.1	Digital high-speed image sequence showing a glottal cycle with 21 tracked glottal contour points per vocal fold: subject JN with normal phonation. Every second frame is shown, the time interval between successively plotted frames is $\Delta t = 0.54$ ms. Note that the right arytenoid cartilage is visible in the left upper corner of the images. Thus, the upper and lower side of each frame correspond to the posterior and anterior side of the vocal folds, respectively. The left and right vocal folds are displayed on the right and left side of the digital images, respectively.	64

LIST OF FIGURES

4.2	Digital high-speed image sequence with 49 tracked glottal contour points per vocal fold of subject WS paralysis (every second frame shown, $\Delta t = 0.54$ ms for plotted frames). The upper and lower side of each frame correspond to the posterior and anterior side of the vocal folds, respectively.	65
4.3	Digital high-speed sequence with 39 tracked glottal contour points per vocal fold of subject MM (every frame shown, $\Delta t = 0.27$ ms). The upper and lower side of each frame correspond to the posterior and anterior side of the vocal folds, respectively.	66
4.4	Selected multi-line kymograms (chosen from 21 kymograms) with highlighted extracted time series for subject JN (with normal phonation). The upper white line corresponds to the left vocal fold edge, the lower white line to the right vocal fold edge. Length of shown kymograms is $T = 81$ ms. The numbers correspond to the scan line number along the posterior-anterior direction. The glottal aperture is black-coded, the surrounding is shown in different gray values. The vertical direction within each kymogram corresponds to the left-right direction of the digital high-speed image frames.	67
4.5	Selected multi-line kymograms for subject WS. Length of shown kymograms is $T = 81$ ms. The upper white line follows the left vocal fold edge, the lower white line follows the right vocal fold edge. In each kymogram desynchronization of the left and right vocal fold oscillation is observed: Five oscillation maxima of the left vocal fold (upper line) correspond to four oscillation maxima of the right vocal fold (lower line).	68
4.6	Selected multi-line kymograms for subject MM. Length of shown kymograms is $T = 67$ ms. The upper white line indicates the left vocal fold edge, the lower white line indicates the right vocal fold edge. In each kymogram nearly symmetrical oscillations of both vocal folds are observed. However, an increasing modulation of both vocal fold oscillations can be seen comparing kymograms 13, 21 and 29 along the posterior-anterior direction. Thus, desynchronization of vocal fold oscillations along the anterior-posterior direction is observed.	69
4.7	Coordinate system for analysis of digital high-speed images: Vocal fold edge points of left and right vocal fold are highlighted. The distances of the estimated glottal midline to the vocal fold edge points are used as vocal fold displacements for further analysis. . . .	71

LIST OF FIGURES

- 4.8 Spatio-temporal plots of the scaled oscillatory component $\delta_k^{(\alpha)}(t_i)$ of vocal fold edge points for subject WS (with left recurrent nerve paralysis). Left and right vocal folds are desynchronized which suggests left-right asymmetry of vocal fold properties. Bright regions display glottal opening, dark regions code for vocal fold edge points with glottal closure. The maximum positive excursions are rescaled to unity, the minimum negative excursions are rescaled to zero. The lower part/upper part of each plot corresponds to the temporal evolution of posterior/anterior vocal fold edge points. During four oscillation maxima on the right vocal fold five oscillation maxima appear on the left side. Vertical homogeneity of the vertical bright stripes during glottal opening indicates no relevant anterior-posterior asymmetry. 74
- 4.9 Spatio-temporal plots of the scaled elongations $\delta_k^{(\alpha)}(t_i)$ for subject MM (with functional dysphonia). During seven oscillation maxima in the upper part of the spatio-temporal plots, six maxima appear in the lower part. The symmetry between left and right spatio-temporal plot is roughly preserved. 75
- 4.10 Comparison of normalized spectra of $\delta_k^{(\alpha)}(t_i)$ of left and right vocal fold at anterior (kymogram line number $k = 40$) and posterior side (kymogram line number $k = 10$) of each vocal fold (normalization to amplitude maximum of all shown spectra). All amplitude maxima can be explained by linear superposition of two independent frequencies $f_r \approx 197$ Hz and $f_l \approx 265$ Hz : $f_{(m,n)} = mf_r + nf_l$. Left-right asymmetry is $Q_{lr}^{(exp)} \approx 0.74$. The dominant peak frequency for both spectra of the right vocal fold is at f_r , whereas for the left vocal fold both spectra have dominant peak frequencies at f_l . The spectral peaks are marked by their corresponding numbers (m, n) . . . 76
- 4.11 Comparison of normalized spectra of displacements $\delta_k^{(\alpha)}(t_i)$ of left and right vocal folds at anterior (kymogram line number $k = 30$) and posterior side (kymogram line number $k = 10$) of each vocal fold (normalization as above): Maxima can be interpreted by linear superposition (m, n) of independent frequencies $f_p \approx 271$ Hz and $f_a \approx 338$ Hz. Anterior-posterior asymmetry is $Q_{ap}^{(exp)} \approx 0.80$ 77
- 4.12 First 21 values of the relative EOF weights for three phonation examples for left (upper row) and right (lower row) vocal folds. The information entropy $S_{tot}^{(\alpha)}$ measures the overall spatial irregularity. As explained in the text, the term “LR biphonation” is introduced for the LR asymmetrical phonation of subject WS and “AP biphonation” for AP asymmetrical phonation of subject MM. 79

LIST OF FIGURES

4.13	First five normalized empirical orthogonal functions for centered time series from subject JN (with normal healthy phonation): Maximum and minimum excursion of the EOFs are shown (normalization to amplitude maxima). The relative weights, indicated in the plots, reflect the contribution of the EOFs to the reconstruction of the observed glottal dynamics. Thus, over 90% of the observed time series is explained by the first EOFs.	81
4.14	First five normalized EOFs for centered time series from subject WS: Maximum and minimum excursion of the EOFs are shown (normalization as above). The first EOFs are enough to explain more than 90% of the time series. Due to scaling the phase shift between anterior and posterior side of the second eof of the right fold can not be seen as clear as on the left fold.	82
4.15	First five normalized empirical modes for centered time series from subject MM: Maximum and minimum excursion of the EOFs are shown (normalization as above). The first two EOFs are required to capture more than 90% of the time series.	83
4.16	Normalized magnitude spectra (linear scale) of first five temporal eigenfunctions for subject JN (with normal healthy phonation) (normalization to amplitude maximum of all shown spectra): All spectra consist of harmonically related peak frequencies together with noise contributions increasing with EOF number.	85
4.17	Normalized magnitude spectra of first five temporal eigenfunctions for subject WS (normalization as above): The left-right asymmetry ratio $Q = f_r/f_l = 0.74 \approx 4/5$ observed in the first temporal eigenfunctions associated with the first EOF reflects the laryngeal asymmetry. The main spectral information is already contained in the first temporal eigenfunctions.	86
4.18	Normalized magnitude spectra of first five temporal eigenfunctions for subject MM (normalization as above): The left-right asymmetry ratio $Q = 1$ of the first two temporal eigenfunctions indicates laryngeal left-right symmetry. AP asymmetry is indicated by the qualitative difference in the spectral content of the first and the second temporal eigenfunctions. Therefore, the main spectral information is contained in the first two EOFs.	87
5.1	Cross section of vocal fold with attached vocal membrane of bat <i>Eptesicus fuscus</i> , bar equals $50\ \mu\text{m}$ (from Suthers and Fattu, 1973) .	97
5.2	Schematic drawing of squirrel monkey larynx (<i>Saimiri sciureus</i>) showing vocal folds with vocal membranes (after Starck and Schneider (1960))	100

LIST OF FIGURES

5.3	Overview of vocalization repertoire of squirrel monkey <i>Saimiri sciureus</i> , grouped into “Peep calls”, “Twit calls”, “Ha calls”, “Arr calls”, “Shriek calls”, and “Combined calls” (Winter et al., 1966)	102
5.4	Experimental setup for brainstem stimulated squirrel monkey phonation: Vocalizations were elicited by neural stimuli of phonatory motor regions. With a flexible angioscope, the vocal folds were observed and recorded with a digital high speed camera. Frame rates of a high speed imaging system of up to 2000 Hz could be used with the available light system. (Fitch et al., 2001)	104
5.5	Comparison of high speed kymogram (time-space plot for single lateral line across squirrel monkey glottis) with acoustical spectrogram (time-frequency plot) revealing tissue vibrations during high frequency phonation instances and arytenoid rocking modulating the lateral width of the larynx (vocal folds, ventricular folds) thus modulating the fundamental frequency: Segment A: prephonatory closing; Segment B: Arytenoid rocking modulating with of larynx corresponds to frequency modulations in spectrogram; Segment C: After opening of glottis short instances of very high frequency phonation, where the limited temporal resolution of the high speed video camera results in blurred instances in the high speed kymograms. (Fitch et al., 2001)	105
5.6	Brainstem stimulated vocalizations of a squirrel monkey: rapid transitions from harmonically structured high frequency call segments either to broadband noiselike behavior or to low frequency harmonic call segments (pitch jump, register jump)	106
5.7	Schematic drawing of nonhuman mammalian larynx as viewed through the thyroid cartilage in the front of the larynx. The thyroid cartilage, positioned on the cricoid cartilage, can rotate about a pivot with the cricoid cartilage. Intrinsic muscles, such as the “vocalis muscle”, the thyro-arytenoid (TA), and the crico-thyroid muscle (CT) determine the geometrical configuration of the larynx. The vocal folds consist of the vocalis muscle and its membranous cover. The vocal membranes are membranous extensions of the cover tissue into the vocal ventricle. Air from the lungs flowing from the trachea passes the glottis formed by the vocal folds and the vocal membrane. During oscillation, air flow pulses (acoustical waves combined with mean flow) are fed upstream into the trachea and downstream into the ventricle, the pharynx and the upper vocal tract. (from Mergell et al. (1999))	110

LIST OF FIGURES

5.8	Sketch of mechanical model for vocal membrane simulations: Vocal folds are represented by box-like masses (m_1, m_2) with elastic (k_1, k_2, k_c) and viscous (r_1, r_2) mechanical restoring forces. Vocal membranes are modeled by reed-like plates (with line mass density m_3/d_3) that oscillate about a pivot on top of the upper vocal fold mass. The pivot incorporates elastic (linear elastic force $k_3\theta$ for excursions θ from the resting position θ_0) and viscous (r_3) mechanical restoring forces. The driving air flow, from the trachea upstream of the vocal membrane model to the vocal tract downstream of the model masses and plates, can induce self-sustained oscillations of the model.	112
5.9	Aerodynamic accelerations for different glottal configurations. Vocal membrane parameters were set at $m_3 = 0.25$ and $d_3 = 0.05$, otherwise standard parameter values were used.	116
5.10	Aerodynamic accelerations for different vocal membrane heights. Vocal folds were chosen parallel: $a_1 = a_2 = 0.05$. Vocal membrane parameters were set at $m_3 = 0.25$ and $d_3 = 0.05$, otherwise standard parameter values were used.	118
5.11	Variation of eigenfrequencies and relative damping times with mass ratio q_m where $d_3 = 0.05$ cm and $\theta_0 = 0$ degree, otherwise standard parameter values were used.	125
5.12	Typical shape of eigenmodes of the vocal membrane model	126
5.13	Variation of eigenfrequencies and relative damping times with resting angle θ_0 [degrees] where $d_3 = 0.05$ cm and $q_m = 2.0$, otherwise standard parameter values were used.	128
5.14	Eigenvectors of the vocal membrane model for $d_3 = 0.05$ cm, $q_m = 2.0$, $\theta_0 = 0$ degree, otherwise standard parameter values. Note that due to the definition of the vocal membrane angle θ , a zero phase difference $\varphi_{A_\theta} - \varphi_{A_2} = 0$ between the upper mass and the vocal membrane corresponds to, e.g., an increase of the glottal area a_2 and a decrease of the vocal membrane area a_{vm}	129
5.15	Eigenvectors of the vocal membrane model for $d_3 = 0.50$ cm, $q_m = 2.0$, $\theta_0 = 0$ degree. Otherwise standard parameter values were used.	130
5.16	Voice onset pressure p_s and Hopf frequency at voice onset as a function of mass ratio q_m and resting angle θ_0 . At voice onset, the fixed point (FP) becomes unstable and a limit cycle oscillation (LC) becomes stable ($q = 1.0$, $d_3 = 0.05$ cm, otherwise standard parameter values are used.).	133

LIST OF FIGURES

5.17	Influence of vocal membrane geometry on phonation onset. For different subglottal driving pressures p_s , phonation onset curves $d_3(\theta_0)$ and their corresponding Hopf frequencies are shown ($q_m = 500.0$, $q = 1.0$, otherwise standard parameter values are used.).	134
5.18	Range of low frequency oscillations varying with tuning parameter q and different mass ratios q_m . The corresponding Hopf frequencies are shown in the right part ($d_3 = 0.05$ cm, $\theta_0 = 1.0$ degree, otherwise standard parameter values are used.).	136
5.19	Onset pressure p_s for low and high frequency registers depending on resting angle θ_0 . The onset curve $p_s(\theta_0)$ reveals a minimum onset pressure for the low frequency register at approx. $\theta_0 = 0.015$ degrees. Note that the corresponding Hopf frequencies are shown in a semi-logarithmic plot ($q = 5.0$, $q_m = 5.0$, $d_3 = 0.1$ cm).	137
5.20	Onset pressure p_s for low and high frequency register as a function of tuning parameter q and vocal membrane height d_3 . The corresponding Hopf frequencies for the onset curves $p_s(q)$ show the register frequencies as functions of the tuning parameter q ($q_m = 10.0$, $\theta_0 = 1.0$ degree).	139
5.21	Simulated time series showing low frequency oscillation of the vocal fold masses and transient damped oscillations of the vocal membrane (The tuning parameter was set to $q = 1.0$, otherwise standard parameter values are used.).	141
5.22	Simulated time series showing high frequency oscillation of the vocal membrane and the vocal fold masses (The tuning parameter is $q = 10.0$, otherwise standard parameter values are used.).	141
5.23	Spectral bifurcation diagram of $\theta_a(t)$ for upward gliding tuning parameter variation (linear tuning parameter variation $q(t)$, $1 \leq q \leq 10$, $p_s = 8.0$ cmH ₂ O, $\theta_0 = 10.0$ degrees, $q_m = 10.0$, $d_3 = 0.05$ cm, otherwise standard parameter values, simulation time: 2000 ms)	144
5.24	Poincare section $\dot{\theta}(t) = 0$ showing subsequent maxima of $\theta_a(t)$ for upward gliding tuning parameter variation (linear tuning parameter variation $q(t)$, $1 \leq q \leq 10$, $p_s = 8.0$ cmH ₂ O, $\theta_0 = 10.0$ degrees, $q_m = 10.0$, $d_3 = 0.05$ cm, otherwise standard parameter values, simulation time: 2000 ms)	145
5.25	Spectral bifurcation diagram of $\theta_a(t)$ for downward gliding subglottal pressure variation (linear pressure variation $p_s(t)$, 37.0 cmH ₂ O $\geq p_s \geq 10.0$ cmH ₂ O, $q = 7$, $\theta_0 = 1.0$ degree, $q_m = 10.0$, $d_3 = 0.05$ cm, otherwise standard parameter values, simulation time: 3000 ms)	147

LIST OF FIGURES

5.26	Poincare section $\dot{\theta}(t) = 0$ showing subsequent maxima of $\theta_a(t)$ for downward gliding subglottal pressure variation (linear pressure variation $p_s(t)$, $37.0 \text{ cmH}_2\text{O} \geq p_s \geq 10.0 \text{ cmH}_2\text{O}$, $q = 7$, $\theta_0 = 1.0$ degree, $q_m = 10.0$, $d_3 = 0.05 \text{ cm}$, otherwise standard parameter values, simulation time: 3000 ms)	148
C.1	Mechanical equivalent of symmetrical vocal membrane model (only left side shown)	194

List of Tables

2.1	Set of standard parameters for the simplified symmetric two-mass model simulating normal male phonation (in the unit system cm, ms, g) (Steinecke and Herzel, 1995)	37
4.1	Details on time series from different subjects with normal and irregular phonation. The sampling rate was 3704 frames per second for all time series. The difference in length is due to the different length of high-speed sequences obtained from the clinical investigation and due to the selection of stationary segments. With respect to the typical oscillation frequencies the number of measured cycles are sufficient for a statistical analysis such as the EOF estimation calculation.	70
4.2	Cumulative sum $\sum \lambda_m^{2(l/r)}$ for the first five values of the relative EOF weights for three phonation types. For subject JN (with normal phonation) and for subject WS (with LR asymmetrical phonation) the first mode already covers more than 90% of the glottal contour dynamics. However, for the subject MM (with AP asymmetrical phonation) the first mode just carries about 80% of the observed time series. Weights of higher modes become rapidly smaller, but are still specific for the type of phonation. To explain more than 97% of the observed glottal dynamics, two modes have to be taken into account for subject JN, two or three modes, respectively, for subject WS; and four modes for subject MM (indicated by numbers in bold face).	80

LIST OF TABLES

- 4.3 Summary of measures quantifying spatio-temporal vocal fold vibration patterns for three phonation examples: Phonation patterns are described by the left-right asymmetry, anterior-posterior asymmetry, the overall spatial irregularity, and the number of dynamically relevant EOF modes. The number of relevant modes is related to the spatial irregularity. Thus it could be connected to the minimum number of degrees of freedom a prospective biomechanical model for the individual pathology should provide. 88
- 5.1 Set of standard parameters for the vocal membrane model adopted from the set of standard parameters for the simplified two-mass model for normal male human phonation (in the unit system cm, ms, g, degree) 122

Chapter 1

Introduction

Acoustical communication in humans and nonhuman mammals is based on the production and perception of voice signals. The primary voice signal (phonation) is generated in the mammalian larynx. Phonation is due to vibratory structures (e.g., vocal folds, ventricular folds, epilarynx) in the larynx. In human speech, vowels are produced by filtering the primary sound signal in the respiratory cavities, the vocal tract, and oral and nasal cavities. Human speech results from a complex interaction between phonation and articulation. Articulation is the variation of the geometry of the filter cavities in the vocal tract (the pharynx, the oral and nasal cavities). In contrast, in nonhuman mammals articulation is less important in communication. Information coding is mainly achieved by the high diversity of the primary sound signal of nonhuman mammals. These complex vocalization patterns are consistently used in specific behavioral contexts.

In human speech, high diversity of the source signal is counterproductive for normal communication. Voice disorders are associated with vocalizations outside a narrow range of primary sound source behaviors. They are in most cases related to pathological changes of the vocal folds. Clinical research on communication disorders includes voice disorders, speech pathologies, and hearing disorders. For clinical diagnosis, voice disorders are mostly classified by perceptual measures of the acoustical speech signal. Objective methods have been proposed to relate the speech signal to classes of pathologies of the vocal folds (Michaelis et al., 1997; Behrman and Agresti, 1998; Blomgren et al., 1998; de Krom, 1993; Giovanni et al., 1999b; Hartl et al., 2001; Muta and Baer, 1988; Schoentgen, 2001; Gerratt and Kreiman, 2000). All these perceptual methods are incapable of reliably separating speech signal into the primary voice signal and contributions of articulation and turbulence to the overall vocal output. This implicates that the underlying mechanisms of pathological voice production and the related speech signal are still hardly understood.

In contrast to speech, where voice disorders are to be avoided, singers (“vocal

1. Introduction

improvisors”) of contemporary vocal music try to enlarge the narrow repertoire of speech signals for artistic purposes. In the search for extended forms of expressions in singing, composers and singers try to explore the production and organization of non-standard vocal music (Griffiths, 1981). For this purpose, traditional descriptions from speech research and linguistics are used (Edgerton et al., 1999; Edgerton, 2002). Similar to the classification of pathological speech, these concepts fail to distinguish clearly between phonation and articulation.

Diversity in vocalizations of nonhuman mammals plays a major role in mammalian communication (Facchini et al., 2003; Laje and Mindlin, 2002; Suthers et al., 1996; Suthers and Margoliash, 2002). Bioacoustic research thus far has mainly focused on documenting the call repertoire, the control of vocalization patterns by the central nervous system and the behavioral context. Since the mammalian larynx was highly conserved during the evolution of different species, the concepts of human voice and human speech research have been applied to animal vocalizations. The study of nonhuman mammalian communication impacts questions on the evolution of human speech and communication. However, detailed understanding of underlying voice production mechanisms in nonhuman mammals is still in its infancy.

The understanding of human phonation is based on the myoelastic-aerodynamic theory by van den Berg (1957). The vocal folds are considered as a visco-elastic layered three-dimensional tissue structure with nonlinear tissue properties. Phonation is generated by self-sustained oscillations induced by aerodynamic forces from the airflow through the larynx. Normal healthy phonation is due to synchronized oscillations of the set of oscillatory modes that is excited by the airflow and coupled due to nonlinearities in the myoelastic-aerodynamic vocal apparatus. Frequently observed voice instabilities occur due to desynchronization of the coupled vocal fold modes. Voice instabilities include subharmonic vocalizations, biphonation (two independent fundamental frequencies), deterministic chaos, and pitch jumps. Examples can be found in newborn cries (Mende et al., 1990), infant vocalizations (Robb and Saxman, 1988), Russian lament (Mazo et al., 1995), normal conversational speech (Dolansky and Tjernlund, 1968), pathological voices (Herzel et al., 1995; Mergell and Herzel, 1997b; Mergell et al., 2000), normal singing (Ward et al., 1969; Gibiat and Castellengo, 2000), human laughter (Bachorowski et al., 2001), contemporary vocal music (Edgerton et al., 2003), and in nonhuman mammalian vocalizations (Wilden et al., 1998; Fitch et al., 2002).

Complex vocalization patterns can be classified and quantitatively understood using the theory of nonlinear dynamics (Facchini et al., 2003; Titze et al., 1993; Fee et al., 1998; Herzel and Knudsen, 1995; Herzel et al., 1995; Jiang et al., 2001b; Jiang and Zhang, 2002a,b). Attractors such as steady state, limit cycle, torus, and chaotic attractor can be related to stationary signals such as prephonatory

1. Introduction

standstill, normal phonation, subharmonic oscillation, biphonation and irregular noise-like voice behavior. Bifurcations such as Hopf bifurcation, period doubling, secondary Hopf bifurcation, and abrupt onset of chaos are induced by varying parameters of the phonatory system. In particular, the theory of coupled oscillators (Bergé et al., 1984) based on the generic circle map can be applied to qualitatively understand the dynamics of the set of coupled oscillators of the vocal apparatus. Nonlinear coupling of these oscillatory modes with, in general, different vibration frequencies leads to frequency locking (mode entrainment), toroidal oscillation, and chaos.

These concepts of nonlinear dynamics can be directly applied to analyze mathematical aerodynamical-biomechanical models of the voice source. The first models consisted of very simple damped mass-spring systems aiming at the synthesis of natural sounding speech (Flanagan and Landgraf, 1968; Flanagan and Cherry, 1969; Ishizaka and Flanagan, 1972). Since then, more and more complex models have been proposed, ranging from multi-mass models (Titze, 1973, 1974) to Finite Element Method models (Alipour-Haghighi et al., 2000). Modeling voice production helped to understand the principles of normal healthy voice production (Titze, 1976; Lucero, 1993, 1999; Berry et al., 1994; de Oliveira Rosa et al., 2003). In recent years, biomechanical models have been developed to describe and understand pathological voices quantitatively (Ishizaka and Isshiki, 1976; Story and Titze, 1995; Mergell and Herzel, 1997b; Mergell et al., 2000; Drioli and Avanzini, 2002; Jiang et al., 2001b; Jiang and Zhang, 2002b). So far, very few studies have been aimed at modeling and reproducing vocalization patterns crucial for animal communication (Fletcher, 1988; Fee et al., 1998; Mergell et al., 1999; Gardner et al., 2001; Smyth and Smith, 2002; Laje and Mindlin, 2002).

The main focus of the present work is to understand basic mechanisms of phonatory instabilities in humans and nonhuman mammals. Therefore, complex spatio-temporal oscillation patterns in pathological phonation will be analyzed quantitatively. Furthermore, voice instabilities in singers associated with contemporary vocal music will be classified using methods from nonlinear dynamics. Finally, animal communication in nonhuman mammals will be studied with a biomechanical model developed especially for nonhuman mammalian larynges.

An introduction to the basic anatomy and physiology of mammalian larynges is given in Chapter 2. For that purpose, the human larynx is used as a representative, as the gross anatomy of mammalian larynges shows little variation between species among mammals. The mechanism of sound production based on the myo-elastic aerodynamic theory of phonation (van den Berg, 1957) will be reviewed. It will be shown that two vibratory modes of the vocal fold tissue are crucial for phonation. The myo-elastic aerodynamic theory implies that phonation can be regarded as a biomechanical-aerodynamical system. Neural control of the configuration of the

1. Introduction

larynx can be regarded as slowly varying parameters compared to phonatory oscillations. An introduction to the theory of nonlinear dynamics will be given and will be related to phonation. Examples of voice instabilities from contemporary singers will be shown with spectral bifurcation diagrams that reveal subharmonic oscillations, biphonation and deterministic chaos of the phonatory signal. A short introduction to the theory of coupled oscillators will be given; It can be applied to the phonatory apparatus when regarded as multiple coupled nonlinear oscillators. Based on the generic circle map, the behavior of coupled oscillators can be classified with two-dimensional bifurcation diagrams. In these bifurcation diagrams, frequency locking and entrained oscillations are revealed by Arnold tongues – regions of nearly rational ratios of the frequencies of two oscillators. Outside of the Arnold tongues, the circle map generated toroidal oscillations. For large coupling, coexisting states and chaos in the circle map are possible. Experimental data on human and nonhuman larynges are obtained with a large variety of methods. An overview of several experimental examination methods of *in vivo* and *in vitro* mammalian larynges will be given. These data can be used as input to mathematical models of mammalian phonation. Several biomechanical models for voice production in humans and nonhuman mammals will be given. In particular, a simplified two-mass model (Steinecke and Herzel, 1995) will be described in detail. This model will be the basis for a new phenomenological model for nonhuman mammalian phonation developed in Chapter 5.

In Chapter 3, complex and multiphonic voice signals of vocal improvisors – singers associated with contemporary vocal music – will be analyzed within the framework of nonlinear dynamics. For this purpose, narrow-band spectrograms will be used as spectral bifurcation diagrams. It will be shown that nonlinear phenomena such as period doubling bifurcations, subharmonic oscillations, biphonation, triphonation, and irregular chaotic phonation is frequent in vocalizations used in contemporary vocal music. Spectral bifurcation diagrams will reveal formant induced bifurcations to subharmonic oscillations and biphonation. The recurrent use of nonlinear phenomena in contemporary vocal music will be demonstrated. Possible production mechanisms such as source-tract coupling and vocal fold desynchronization will be discussed. These mechanisms allow vocal improvisors to specifically use nonlinear phenomena in a reproducible way for musical tasks.

In Chapter 4, spatio-temporal patterns of high speed glottograms (HGGs) will be analyzed. HGGs allow time-resolved observation of the superior vocal fold edge during sustained phonation. The extracted time series of points along the superior vocal fold edge will be decomposed into empirical eigenmodes. HGGs of pathological phonation showing biphonation will be analyzed and compared to healthy normal phonation. It will be shown that in pathological biphonic phonation, the number of excited modes is significantly higher than in normal phonation. This

1. Introduction

analysis shows a new mechanism for glottal biphonation found *in vivo*: Biphonation will be shown to be induced by anterior-posterior asymmetry within a single vocal fold. So far, biphonation was observed to be induced by asymmetries between the left and right vocal fold. Empirical eigenmode analysis will be shown to be an appropriate tool to quantify glottal left-right and anterior-posterior asymmetries observed in high speed recordings of oscillation vocal folds. Furthermore, spatial irregularities of the observed glottal dynamics will be described with an entropy measure. This clearly shows that in pathological phonation, the complexity of excited spatio-temporal modes is significantly higher than in normal phonation. Finally, production mechanisms inducing anterior-posterior desynchronization of vocal fold modes and the clinical significance of the used analysis method will be discussed.

In Chapter 5, a phenomenological model for nonhuman mammalian larynges with vocal membranes will be developed and analyzed. It will aim at the description and understanding of vocalization patterns in bats and nonhuman primates. Vocal membranes, frequently found in nonhuman mammalian larynges, consist of lightweight membranous upward extensions of the vocal folds. In bats, vocal membranes are responsible for ultrasonic calls used for prey capture and orientation during flight. In primates, they facilitate a highly diverse call repertoire. *In vivo* and *in vitro* experiments with bats and primates showed that nonlinear phenomena are frequent and are used consistently in specific behavioral contexts. Spectrograms of primate calls reveal subharmonics, biphonation, register jumps, chaotic behavior. Coexisting attractors and abrupt transitions between different oscillatory behaviors are documented in the literature on bats and primates. For modeling, the simple two-mass model developed for human phonation (Steinecke and Herzel, 1995) will be extended by oscillating plates as upward extensions. This new model generalizes a previous model with static reed-like plates (Mergell et al., 1999). In general, modeling nonlinear phenomena in bats and primates is based on adding a third oscillator on top of the vocal folds. This additional oscillator has an eigenfrequency that is orders of magnitude higher than the typical modal frequencies found in vocal folds. The major goal will be to analyze the influence of vocal membranes on the dynamic behavior of nonhuman larynges. Linear eigenmode analysis will reveal the effects of the chosen vocal membrane geometry on eigenfrequency ratios and eigenmode shape. On the background of coupled oscillators, the ratio of eigenfrequencies is important for entrainment of oscillatory modes (frequency locking) in the full nonlinear model. Eigenmode shapes effect the onset behavior of the full model. As a central result, the vocal membrane model will be shown to have two voice registers. These registers could overlap. It will be argued that within the overlapping register region, biphonation can occur. As another important result, it will be shown that the vocal membranes design

1. Introduction

can be optimized with respect to minimal phonation onset pressure and a large phonatory pressure range. Finally, numerical integrations with slowly varying parameters will demonstrate that the vocal membrane model reproduces complex vocalization patterns in bats and primates qualitatively.

In general, in this thesis voice instabilities in human and nonhuman mammals will be analyzed and modeled. Voice instabilities in humans can be caused by laryngeal pathologies, but they can also be induced in healthy normal larynges for artistic purposes. In nonhuman mammals they are crucial for the diversity of vocalization repertoire that is the basis for animal communication. Experimental data will be classified using methods from nonlinear dynamics. Comparison with simulated data from biomechanical models will allow deeper insight into mechanisms for voice instabilities in humans and nonhuman mammals.

Chapter 2

Physics of Voice Production

2.1 Anatomy and physiology of the voice production apparatus

In the following chapters, spatio-temporal signals of the mammalian phonatory system will be analyzed and synthesized. Therefore, a brief introduction to the anatomy and physiology of the voice production apparatus is necessary. Here, a short overview of the basic anatomical components of the human voice apparatus is given. After a description of the human voice production mechanism, the human phonatory system will be embedded into the theory of nonlinear dynamics. Finally, standard models for a qualitative and quantitative simulation of voice production will be shown. For an extended introduction to the anatomy and physiology of the human voice production system see, e.g., Titze (1994a) and Wendler et al. (1996).

Central to the production of voiced speech is the larynx, the housing of the vocal folds. The vocal folds confine the glottal space (glottis), and are located between the respiratory system and the vocal tract (Fig. 2.1). Their primary function is the protection of the trachea and the lungs (respiratory system) from secretions, food, and foreign material during swallowing. Additionally, in many mammals the vocal folds are the primary phonation organ. Acoustically, they serve pressure waves into the subglottal tract (trachea, lungs) and the supraglottal airway (vocal tract). The vocal tract comprises several cavities, most of which can be modified by articulators. The shape of the pharyngeal cavity is manipulated by the pharynx muscles. The shape of the oral cavity can be altered by movement of the tongue, the lips, the mandible, and the velum. The geometry of the nasal cavity is determined by facial muscles. The acoustical coupling of the nasal cavity to adjacent cavities is adjusted, e.g., by the velum.

Anatomically, the skeletal framework of the larynx consists of cartilages, extrinsic and intrinsic muscles, and mucosal tissue as coating material (Fig. 2.2).

2. Physics of Voice Production

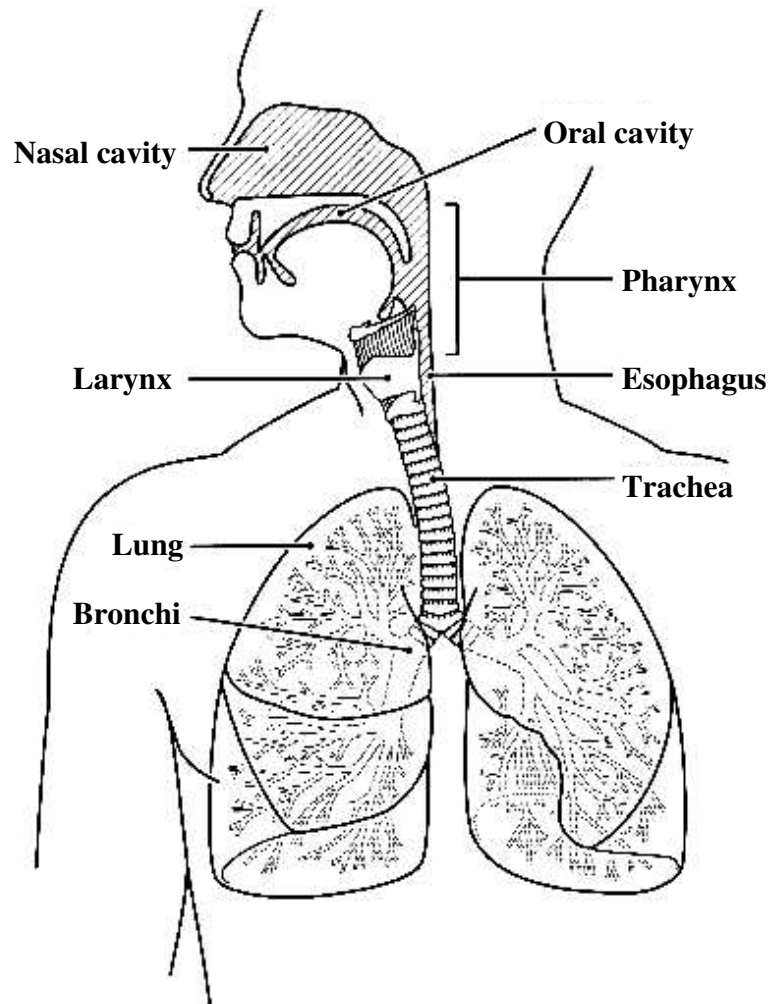


Figure 2.1: Components in the airway system in the head, neck, and chest (Titze, 1994a): The lungs produce pressure that drives the subglottal airstream. The airstream is fed to the larynx via bronchi and trachea. The primary function of the larynx is to protect the airway system from foreign material (such as food) which passes into the esophagus during swallowing. The vocal tract, comprised of pharynx and oral and nasal cavities, filters the primary sound signal generated by airstream-driven oscillations of the vocal folds in the larynx.

2. Physics of Voice Production

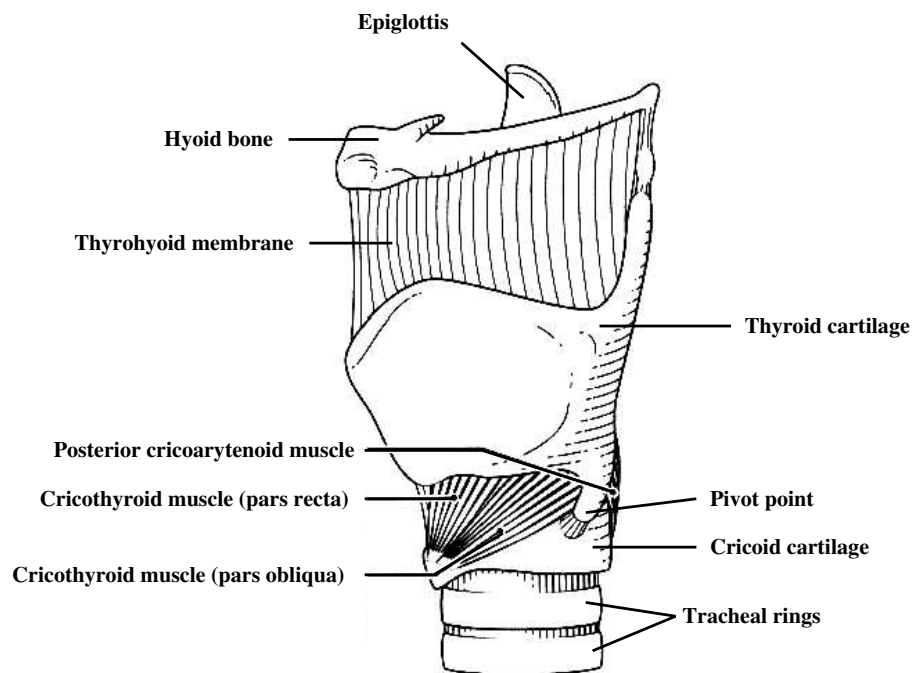


Figure 2.2: Sideview of the skeletal laryngeal framework (Titze, 1994a): Two tracheal rings, the cricoid and the thyroid cartilages, the hyoid bone, and the epiglottis are shown. Ligaments and intrinsic muscles connect the cricoid and thyroid cartilages. Cricothyroid muscle contraction rotates the cricoid cartilage against the thyroid cartilage, which lengthens the vocal folds. The cricoarytenoid muscle contraction rocks (rotates and translates) the arytenoid cartilages to open and close the glottis.

2. Physics of Voice Production

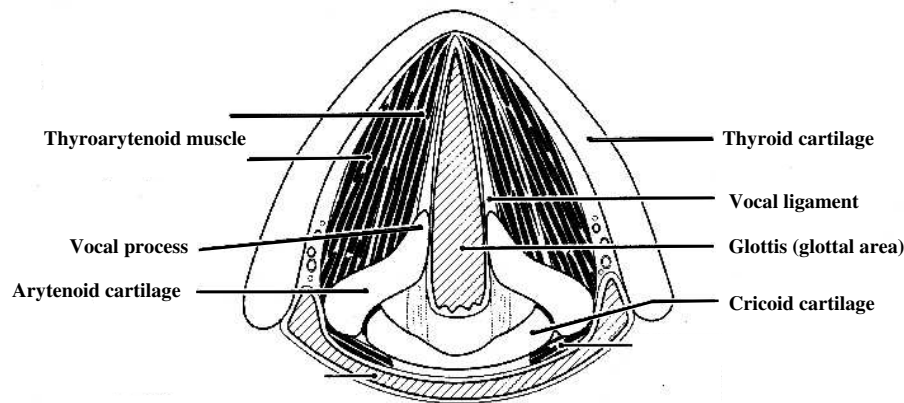


Figure 2.3: Superior view of the muscular and cartilagenous framework of the larynx (transverse section at the level of the vocal folds) (Titze, 1994a): The vocalis muscle (thyroarytenoid muscle) is inserted between the arytenoid and thyroid cartilages. The arytenoid cartilages can rock and move on top of the cricoid cartilage due to contraction of the thyroarytenoid muscle, the cricoarytenoid muscles and the interarytenoid muscle. The top side of the cricoid cartilage supporting the arytenoid cartilages can be moved back and forth by cricothyroid muscle and thyroarytenoid muscle contraction, respectively.

The cricoid, thyroid, and arytenoid cartilages are connected by ligaments and the intrinsic muscles (as agonist-antagonist muscle pairs) that determine the relative position of the cartilages to each other. Therefore, the intrinsic muscles can manipulate the geometry of the glottis. They also adjust the geometry and tissue tension of the vocal folds. The extrinsic muscles determine the position of the larynx with respect to surrounding anatomical structures of the head and the neck. They mainly adjust the geometry and acoustical coupling of the glottal area with adjacent sub- and supraglottal cavities. Innervation of the intrinsic muscles, and thus neural control, is accomplished through two major branches of the vagus nerve (see Fig. 2.5). One branch, the superior laryngeal branch, innervates only the cricothyroid muscle, the muscle between the cricoid and the thyroid cartilage (Fig. 2.2). All other intrinsic laryngeal muscles are innervated by the recurrent laryngeal branch of the vagus nerve.

Morphologically, the human vocal folds can be divided up into several layers (see, e.g., Hirano, 1974): Mucosa, ligament, and vocalis muscle. Some authors subdivide the mucosa into epithelium and superficial layer, and the ligament into intermediate and deep layer (see Fig. 2.4 from Titze et al. (1994)).

From a physicist's point of view, the larynx can be regarded as a valve between

2. Physics of Voice Production

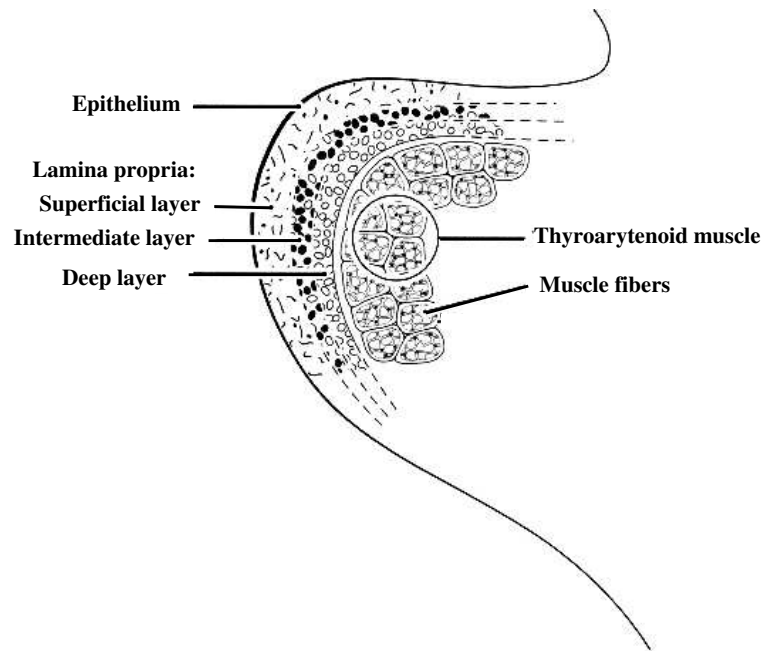


Figure 2.4: Schematic of a frontal section through one vocal fold (Titze, 1994a): The tissue layers can be roughly grouped into vocal fold body and vocal fold cover. The body consists of the muscle fibers of the thyroarytenoid muscle. The cover is built by the epithelium and the lamina propria (superficial, intermediate, and deep layer).

2. Physics of Voice Production

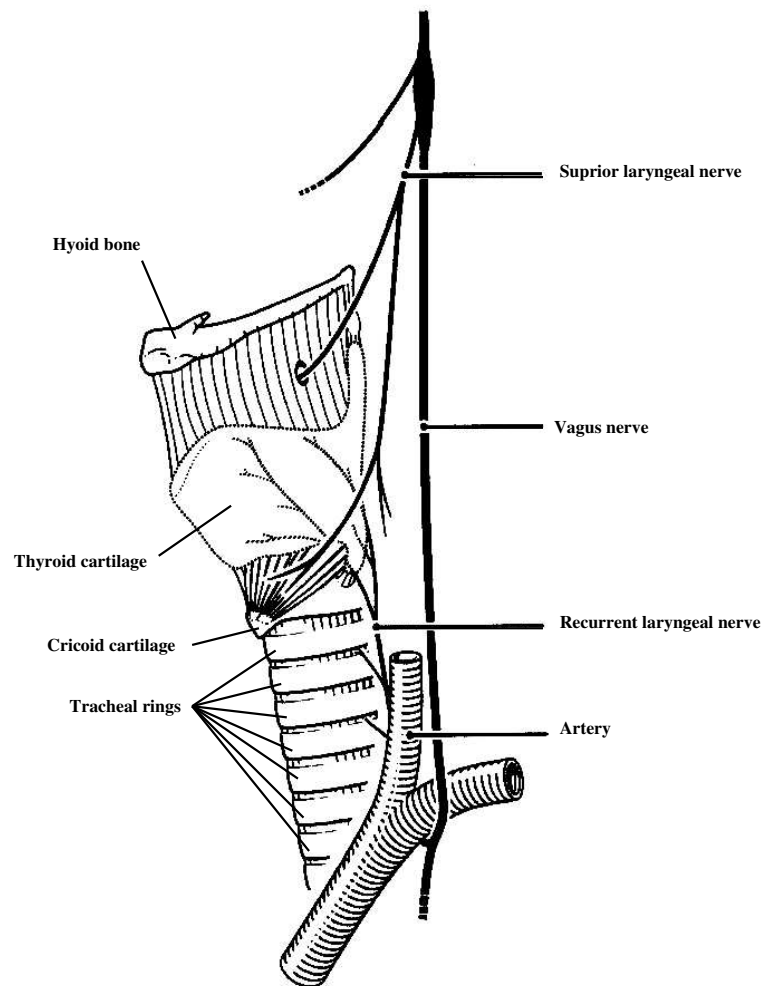


Figure 2.5: Innervation of the larynx (Titze, 1994a): Two branches of the vagus nerve innervate the intrinsic laryngeal muscles: The superior laryngeal nerve only innervates the cricothyroid muscle. The recurrent laryngeal nerve innervates all other intrinsic muscles (including agonistic-antagonistic muscle pairs).

2. Physics of Voice Production

the trachea and the vocal tract. The airflow provided by the pressure in the lungs is gated by the vocal fold valve.

In aeracoustical terms, oscillating vocal folds act as a vibrating orifice that modulate the confined airflow.

The air pulses are then filtered acoustically by the vocal tract ranging from the epiglottis to the mouth and nasal tract. The vocal folds can be regarded as a bilateral, three-dimensional and multilayered anisotropic glottal valve structure. Note that this structure is common to all mammalian larynges (Negus, 1949; Harrison, 1995; Schön-Ybarra, 1995; Fitch and Hauser, 1995). Due to tissue properties (elasticity and viscosity), two gross tissue parts of the vocal folds can be differentiated: The thyroarytenoid muscle can be referred to as the vocal fold body (see Fig. 2.4), being an active myoelastic tissue with anisotropic tissue properties. Due to muscle fibers there is a difference between longitudinal and transversal elasticity and viscosity. Mucosa and ligament can be referred to as the vocal fold cover, a passive tissue with isotropic tissue properties.

2.2 Mechanism of sound production

In the following, sound production will refer to the generation of the primary sound signal in the larynx. Discussing voice production in general, phonation, articulation and speech have to be distinguished. *Phonation* describes the process of the primary sound generation due to air pulses in the larynx. *Articulation* refers to filtering of the primary acoustical signal by the vocal tract determined by the geometry of the pharynx, the oral and nasal cavities, and the lips. *Speech* is the coding of information using voiced and unvoiced sounds in a meaningful temporal sequence (Petursson and Neppert, 1991).

Physically, phonation results from the interplay of a driving fluid (the air) with a complex visco-elastic medium (the vocal folds). Vocal fold oscillations occur due to a flow-induced instability of the visco-elastic vocal fold tissue. The mechanism for self-sustained vocal fold oscillations is termed myoelastic-aerodynamic theory (van den Berg, 1957; Titze, 1980). It consists of an inherent feedback loop: Glottal geometry determines the subglottal, intraglottal and supraglottal airflow. The resulting aerodynamical forces (due to the transglottal pressure distribution and viscous air-tissue interactions) act on the vocal fold surfaces. They determine in turn the glottal geometry. For low fluid flow rates, there is a stable static equilibrium state of the feedback loop. It loses stability at a critical flow rate. Vocal fold oscillations result as a new stable dynamical equilibrium for flow rates above the threshold.

Before self-sustained oscillations set in, activity of adduction muscles (intrinsic muscles) close the glottis. These muscles rotate and translate the arytenoid

2. Physics of Voice Production

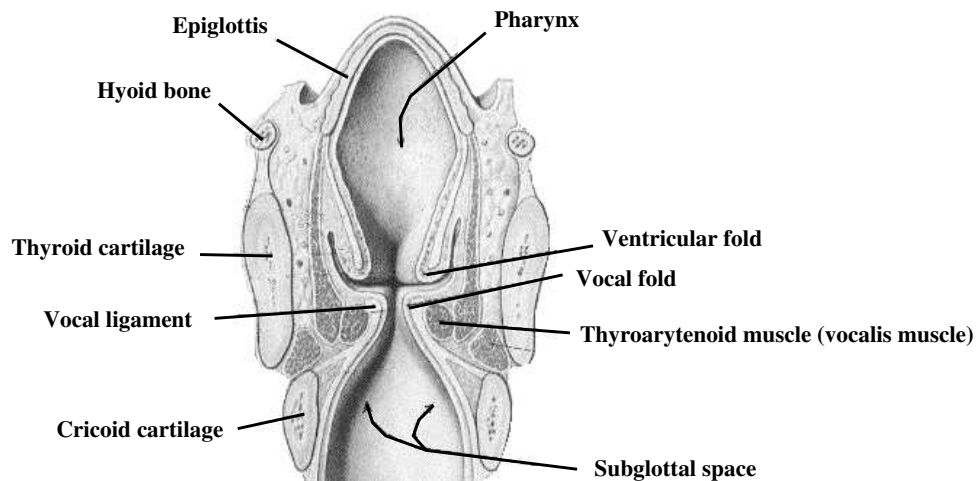


Figure 2.6: Frontal section of the larynx through the center of the glottis: The vocal folds are shown in prephonatory rest position. Adduction muscle activity narrows the glottis (the space between vocal folds), thus increasing subglottal pressure. Aerodynamical forces blow these visco-elastic orifice apart, until elastic tissue recoil and aerodynamical Bernoulli forces close the glottis closes. [The ventricular folds, a valve structure similar to the vocal folds, are shown above the vocal folds. Note that in certain phonatory modes they have been observed to oscillate with the vocal folds (Fuks et al., 1998). Their aerodynamical influence of facilitating phonation in whispered voices has been shown by Miller et al. (1988). de Oliveira Rosa et al. (2003) demonstrated that they could increase the transglottal pressure drop.]

2. Physics of Voice Production

cartilages, they rotate the cricoid cartilage against the thyroid cartilage, and they bulge the vocal folds into the glottal space (Fig. 2.2 and 2.3). The prephonatory narrowing or closing of the glottis corresponds to the prephonatory rest position.

Contraction of the rib cage expels air from the lungs. The airflow through the narrowed glottis or against the fully closed glottis increases intraglottal or subglottal pressure, respectively. After a threshold pressure is reached, the vocal folds are blown apart. The resultant increasing tissue tension in the vocal folds due to deformation produces an elastic recoil. Additionally, the transglottal pressure (pressure difference between outlet and inlet of the vocal folds) drops due to increased airflow in the narrow glottal channel (Bernoulli effect). These restoring forces bring the vocal folds back towards the glottal midline, thus narrowing or closing the glottis again. Then, the next glottal cycle starts. The oscillation frequency is related to the perceived pitch of the voice. The tissue tension for the elastic recoil is the result of both passive tissue elasticity in the mucosal covering layers and the active myoelasticity of the muscular tissue.

In a frontal section (sideview of larynx cut in the middle of the glottis, e.g., Fig. 2.6) a sketch of the glottal oscillation cycle reveals two main vibrating modes (in terms of coherent spatial motion). The spatially complex behavior of the vocal folds (see Fig. 2.7) can be separated into a basic opening and closing mode and a mucosal wave mode. The basic mode, that mainly governs the oscillation of the bulk of the vocal folds (vocal fold body associated with the vocalis muscle), is characterized by an in-phase movement of upper and lower medial edges of the vocal folds. The oscillation of the vocal fold cover tissue results in the mucosal wave mode, where the upper and lower medial edges of the vocal folds vibrate out of phase. The general oscillation pattern of vocal folds is the result of the superposition of these two vibratory modes. Vocal fold vibrations in different voice registers (pulse register (Stroh bass), chest (modal) voice, falsetto, and whistle register) can be described by the particular superposition of these two modes.

Phonation control (oscillation frequency, amplitude, and spectral content) is achieved by the intrinsic muscles, subglottal pressure, and the shape of the vocal tract. The cricothyroid muscle that is innervated by the superior laryngeal branch of the vagus nerve (Fig. 2.5) crucially determines the stiffness and length of the vocal folds. The contraction of this muscle rotates the thyroid cartilage about the cricothyroid joint. This rotation lengthens and stiffens the oscillating tissue suspended between the thyroid cartilage and the arytenoid cartilages (see Fig. 2.3). Fine-tuning of both the myoelastic properties of the vibrating tissue and the shaping of the vocal folds is achieved by the thyroarytenoid muscle (vocalis muscle, the body of the vocal folds, see Fig. 2.3). For certain voice registers (pulse register, modal voice) the thyroarytenoid muscle tissue is itself exposed to large tissue vibrations, thus being both manipulator and oscillator.

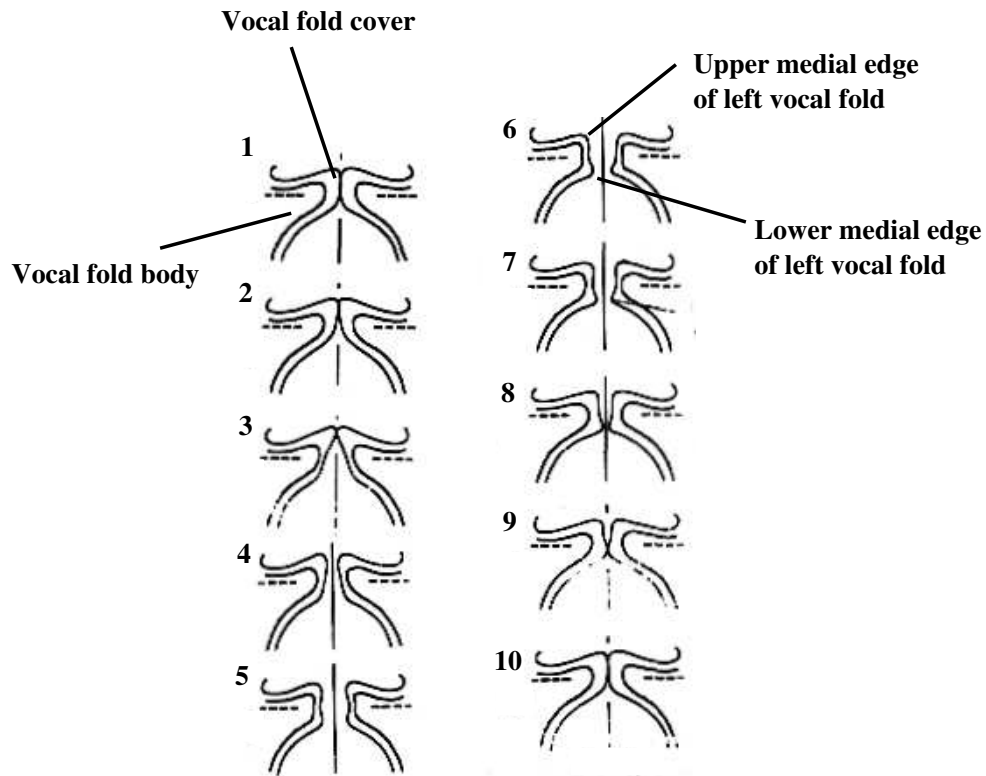


Figure 2.7: Schematic sideview of one glottal cycle: After prephonatory closing of the glottis, rising subglottal pressure opens vocal folds starting from below. The upper edge follows the motion of the lower edge with a certain phase shift. Due to increasing elastic tissue recoil first the lower portion of the vocal folds starts closing again. The upper portion again follows with a certain delay. (after Schönhärl, 1960)

2. Physics of Voice Production

The oscillation amplitude for fixed phonation frequency is mainly controlled by the subglottal pressure provided by lung contraction. The spectral content of the radiated sound from the mouth is described by the overall sound intensity and voice timbre. The sound intensity is mainly adjusted by the shape of the vocal tract, which can be described by an acoustical transfer filter function. Two variables determine voice timbre: the shape of the vocal tract (related to transfer filter function), and the duration of the opening time of the glottis within the glottal oscillation cycle.

The physiological parameters for phonation control (muscular activity, tissue stiffness, vocal fold geometry) are slowly varying parameters. The time scales of neural signals and muscle response for humans are up to approx. 20 Hz (Riviere et al., 1998)). The time scale of human phonation is in the range of 100–1000 Hz for speech and singing voice. This separation of time scales for control parameters and dynamic behavior allows to simplify phonation as a biomechanical-aerodynamical system.

2.3 Nonlinear dynamics of phonation

The physiological mechanisms for sound production in the larynx are known in principle. Given a complete physical description of the properties of the laryngeal passive tissue, the innervation and activity of the muscular tissue, and the interaction with the airstream from the lungs, the spatio-temporal behavior of the sound production apparatus could be determined for given initial conditions. The governing equations of the tissue properties (viscosity and elasticity) and the driving airstream (Navier-Stokes equations), as inherently nonlinear, imply that for given initial conditions the temporal behavior of the overall biomechanical-aerodynamical system is deterministic, and might be unpredictable (chaotic).

Following standard textbooks (e.g., Bergé et al., 1984), the basic principles of (nonlinear) deterministic dynamical systems will be reviewed.

Here, autonomous systems without explicit time dependence of the driving force are discussed. The state vector $\vec{x}(t) \in \mathbb{R}^n$ in phase space, composed by the dynamical variables of the nonlinear dynamical system, changes due to a general force \vec{F} . The general force depends on the current state $\vec{x}(t)$ and on system parameters $\vec{\lambda}$. The dynamics of the state vector $\vec{x}(t)$ can be written as a first order differential equation:

$$\frac{d}{dt}\vec{x}(t) = \vec{F}(\vec{x}, \vec{\lambda}) \quad , \quad \vec{x} \in \mathbb{R}^n \quad , \quad \vec{\lambda} \in \mathbb{R}^m \quad , \quad t \in \mathbb{R} \quad (2.1)$$

The phase flow $\mathbf{f}_{\vec{\lambda}}$ is the associated mapping of initial states $\vec{x}(t=0) =: \vec{x}_0 \in \mathbb{R}^n$

2. Physics of Voice Production

to states $\vec{x}_t^{\vec{x}_0}$ at times t .

$$\vec{x}_t^{\vec{x}_0} = \mathbf{f}_{\vec{\lambda}} \vec{x}_0 \quad (2.2)$$

The phase flow $\mathbf{f}_{\vec{\lambda}}$ depends on the parameters $\vec{\lambda}$, a vector in the parameter space \mathbb{R}^m containing the set of model parameters (e.g., for the phonatory system: subglottal driving pressure of the air, glottal geometry, tissue stiffnesses and viscosities, etc.). Starting from initial states \vec{x}_0 , the phase flow $\mathbf{f}_{\vec{\lambda}}$ generates a set of states $\vec{x}_t^{\vec{x}_0}$ ($t \in \mathbb{R}$), a trajectory in phase space, depending on the associated initial state.

As the phonatory system is an open system dissipating energy to the surrounding environment due to air and tissue viscosity, dynamical systems describing phonation are dissipative systems. Therefore, the volume of initial regions in phase space contracts during time due to the phase flow $\mathbf{f}_{\vec{\lambda}}$ mapping. Contraction of phase space volumes asymptotically generates sets of phase space points called attractors, a bounded attractive set of phase space points. Different disjoint regions of phase space, termed basins of attractor, are mapped to different attractors by the phase flow $\mathbf{f}_{\vec{\lambda}}$. Trajectories starting from an attractor basin approach asymptotically the corresponding attractor. Points already within the attractor set are mapped to points within the attractor, i.e. an attractor is invariant under the phase flow.

Attractor sets can be subdivided into four different classes (e.g., Jackson, 1989): steady state, limit cycle, torus, and chaotic attractor. Steady states are represented by fixed points in phase space. For instance, the trajectories of a damped harmonic oscillator approach asymptotically the stable stationary rest state. Limit cycles are closed curves in phase space (as in, for example, the asymptotic behavior of a driven, damped harmonic oscillator). A third attractor type, termed torus, is a two-dimensional object in phase space, where the trajectory of two coupled oscillators asymptotically moves on the surface, never meets itself again and asymptotically fills the whole surface. A chaotic attractor is typically a fractal set of points in phase space. Stressing the fractal nature of this set, it is termed strange attractor. Stressing sensitive dependence of single trajectories on initial conditions and exponential divergence of initially close trajectories, the term “chaotic attractor” is used (Strogatz, 1994). A single trajectory on a strange attractor is aperiodic, irregular, and deterministic, but not random or stochastic. For instance, even only two oscillators, if strongly coupled, can exhibit chaotic irregular oscillations (Bergé et al., 1984).

For constant control parameters $\vec{\lambda}$ and given initial conditions, the long time behavior of a nonlinear system is uniquely determined by the attractor types associated to the attractor basins. For varying components of the parameter vector $\vec{\lambda}$, the local topology of attractors in phase space can change qualitatively. At critical parameter values, bifurcations between different attractor types are induced that qualitatively change the asymptotic behavior of trajectories. (For an introduction

2. Physics of Voice Production

to the theory of local bifurcations in dynamical systems see, e.g., Jackson, 1989; Guckenheimer and Holmes, 1983). For phonatory systems the most important bifurcation types are: Hopf bifurcation, secondary Hopf bifurcation, period doubling bifurcation, and abrupt onset of deterministic chaos. At a Hopf bifurcation a stable fixed point becomes unstable and a stable limit cycle oscillation governs the long term behavior of trajectories. Secondary Hopf bifurcations are due to an emergence of a second oscillation with an independent (incommensurable) frequency, changing the behavior from a limit cycle oscillation to a toroidal oscillation. At a period doubling bifurcation, an emerging second frequency at half the frequency of a limit cycle oscillation changes a limit cycle attractor to a folded limit cycle attractor. Cascades of period doubling bifurcations or secondary Hopf bifurcations can be precursors to the onset of deterministic chaos (For further routes to chaos see, e.g., Bergé et al., 1984).

Given a complete description and a precise measurement of dynamical variables and parameters of, e.g., the phonatory system, attractor and bifurcation analysis could reveal all possible different dynamical behaviors of the system. Unfortunately, in real-world systems, not all variables are directly measureable and values of system parameters might be inaccurate. Time series analysis of a measured signals provides a first step to understanding the attractors of an experimental dynamical system. Attractors can be classified by attractor dimensions and Lyapunov exponents (Ebeling et al., 1990). From a measured, stationary signal, a pseudo-phase space, diffeomorph to the system attractor, can be reconstructed. The experimental time series is embedded into higher-dimensional spaces (defined by, e.g., delay coordinates, or derivative coordinates). Estimated attractor type, attractor dimension and Lyapunov exponents in pseudo-phase space are equivalent to the true attractor of the system.

For phonation, there is no unique and invertible one-to-one mapping from estimated attractors to laryngeal configurations in terms of parameters λ . This is due to coexisting attractors in phase space, the small number of attractor types, and a high-dimensional parameter space. For instance, various voice pathologies might lead to subharmonics.

In the framework of nonlinear dynamical systems shown above, the physiological mechanism of sound production in the mammalian larynx can be described as follows: The prephonatory standstill of the vocal folds, approaching and making contact with each other to close the air passage between them and to increase subglottal pressure, is a fixed point of the biomechanical-aerodynamical system. Phonation onset to stable sinusoidal vocal fold oscillations is a Hopf bifurcation from a fixed point to a limit cycle oscillation of the biomechanical-aerodynamical system. The two main vibratory modes of left and right vocal fold (vocal fold body mode and mucosal wave, c.f. Fig. 2.7) are oscillating synchronously. In healthy

2. Physics of Voice Production

and in pathological voices, desynchronization of vibrating modes has been observed. For example, under certain laryngeal conditions, left and right vocal folds could lose entrainment (synchronization), which could cause subharmonic oscillations via a period-doubling bifurcation, or toroidal oscillations via a secondary Hopf bifurcation. Toroidal oscillations have been termed *biphonation*, a behavior characterized by two different audible frequencies. Under extreme phonation conditions, even abrupt transitions to chaotic behavior can be observed where the vocal folds oscillate in an irregular aperiodic way.

Rigorous evidence of nonlinear phenomena in voice signals (subharmonic oscillations, biphonation, deterministic chaos) can be provided by sophisticated time series analysis as sketched above. Time-frequency plots, such as spectrograms, sonagrams (visible speech) are spectral bifurcation diagrams where time is a control parameter (Lauterborn and Suchla, 1984; Lauterborn, 1986; Lauterborn and Parlitz, 1988). To illustrate, spectrograms of extra-complex sonorities of vocal improvisors explain the interpretation of spectral bifurcation diagrams (see Figures 2.8, 2.9, 2.10, 2.11). Voice samples from trained vocal improvisors show reproducible, intentional and controlled production of nonlinear phenomena in healthy larynges (see also Chapter 3).

Fig. 2.8 shows limit cycle oscillations after a Hopf bifurcation from a fixed point (voice onset). A stack of harmonics illustrated by equally spaced horizontal lines above the fundamental frequency (here at about 500 Hz) in the spectrogram represents limit cycle vibrations of the vocal folds. At certain times additional parallel lines in between the harmonics appear at multiples of $1/2$, $1/3$ or $1/4$ of the original fundamental frequency. These subharmonic oscillations are induced by transitions between entrainment regions (period doubling, tripling, quadrupling bifurcations) of the phonatory system.

In Fig. 2.9, two close and uncorrelated moving frequency bands (here at about 800 Hz) are visible together with several spectral lines above. These could be explained by linear combinations of the two, uncorrelated moving frequencies. The independent (incommensurable) frequencies are introduced by a secondary Hopf bifurcation from a limit cycle oscillation. The phonatory system exhibits toroidal oscillations.

In Fig. 2.10 sharp transitions from regular stacks of harmonics to broadband noiselike segments spectrally represent sudden onsets of deterministic chaos from limit cycle oscillations. In the broadband chaotic segments, distinct spectral peaks can be observed that are related to the regular stacks of harmonics before the onset of chaos. Within the chaotic regime the phonatory system could oscillate in an almost periodic way due to an unstable limit cycle. Typical for chaotic behavior, transitions to regular limit cycle oscillation and back to chaotic irregular behavior are shown in this example.

2. Physics of Voice Production

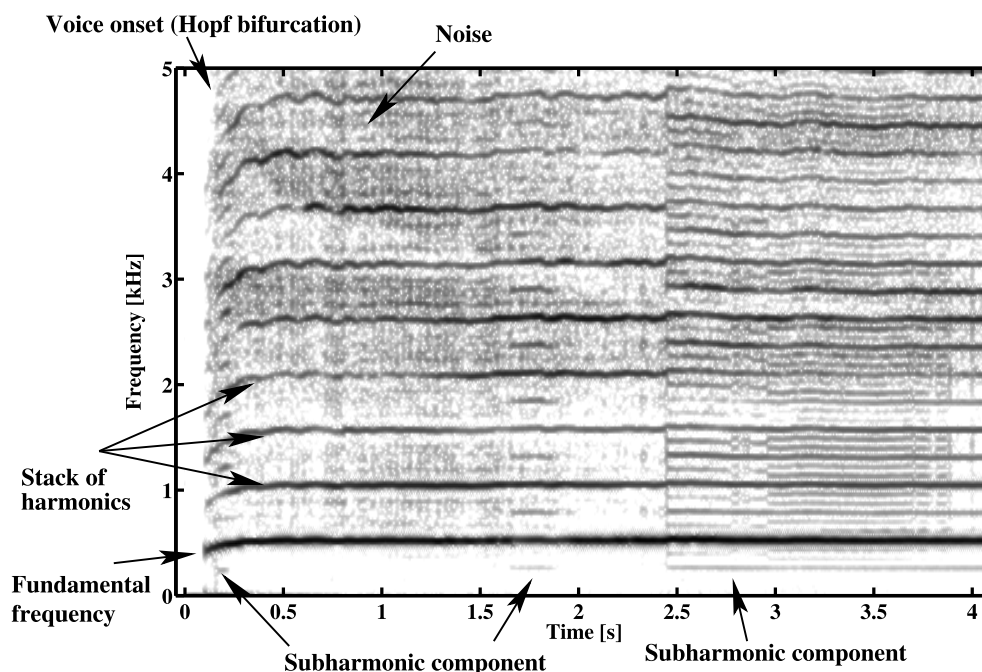


Figure 2.8: Spectrogram of sustained phonation of a trained voice with various subharmonic components: In the above time-frequency plot, the subsequent power spectra calculated from moving time segments are shown. The normalized sound intensity distributions (on a logarithmic scale) are transformed to distributions of gray scale values. Here, black points indicate the highest intensities. This spectrogram shows a segment of sustained phonation after voice onset, induced by a Hopf bifurcation from the prephonatory fixed state. Period doubling bifurcations inducing subharmonic oscillations at small integer ratios of the original fundamental frequency are visualized (see arrows for “Subharmonic components”). As in the above spectrogram, subharmonic oscillations show up as additional parallel bands of high intensity within the stack of harmonic frequencies, integer multiples of the fundamental frequency (here at about 500 Hz). The gray shading between the harmonics and subharmonics is due to noise components.

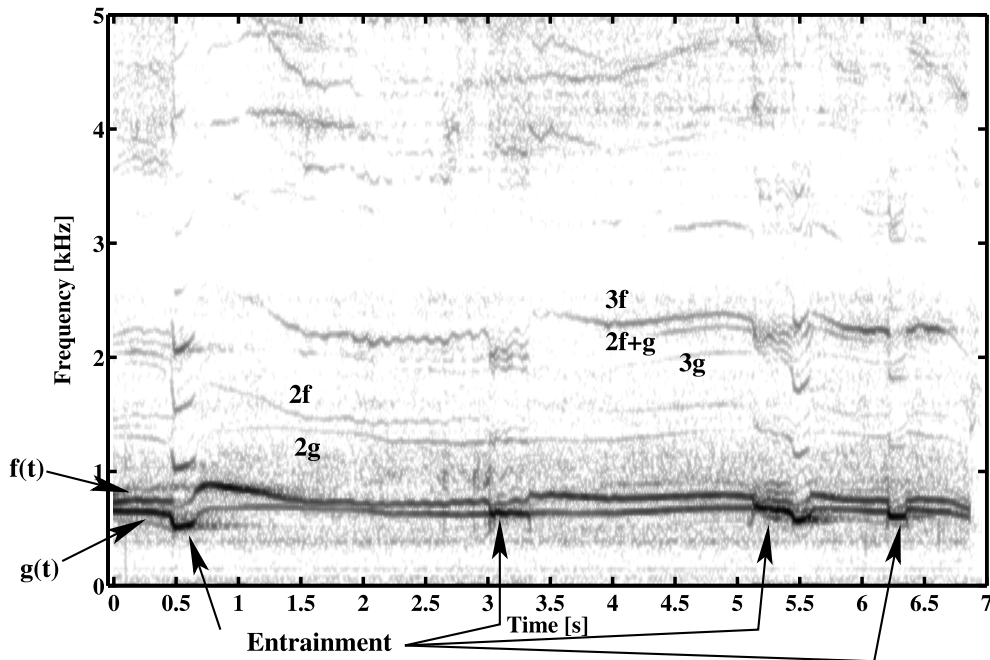


Figure 2.9: Spectrogram of sustained phonation revealing simultaneous unrelated pitch melodies $f(t)$ and $g(t)$ in the voice of a single singer. Toroidal oscillations of the singer's phonatory system are shown: Two frequency bands with high intensities (here at about 800 Hz) vary uncorrelated over time. Short segments of entrained oscillation, associated with folded limit cycle oscillation, are interspersed: During short time segments the uncorrelated behavior of the pitch melodies $f(t)$ and $g(t)$ abruptly changes to entrained oscillations (e.g., at 0.5 sec). Frequency components at linear combinations of f and g appear in the spectrogram due to nonlinear coupling of the two independent oscillations.

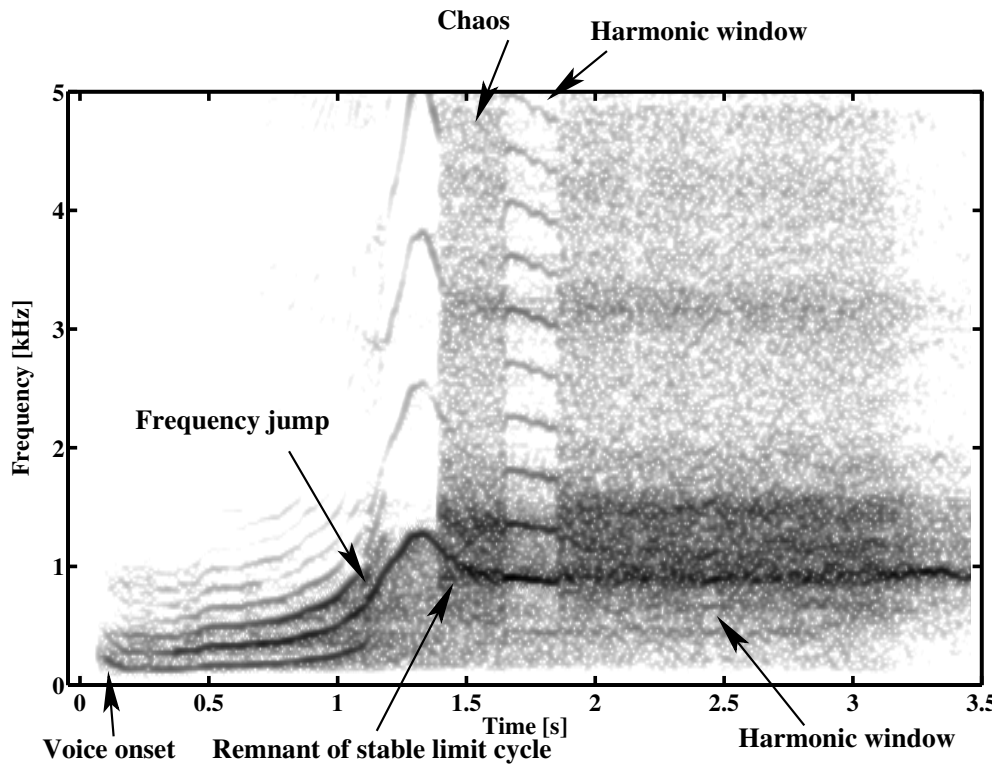


Figure 2.10: Spectrogram of sustained phonation of the voice of a singer showing an abrupt onset of chaotic irregular oscillations: Following voice onset, a frequency jump at about 1.1 sec occurs during the increase of the fundamental frequency. The harmonic oscillation behavior changes to irregular noise-like behavior. Embedded in the broadband chaotic segment, a remnant of the previous fundamental frequency is visible. This is due to the presence of the previously stable limit cycle oscillation. The above example shows harmonic windows with folded limit cycle oscillations interspersed in the chaotic segment.

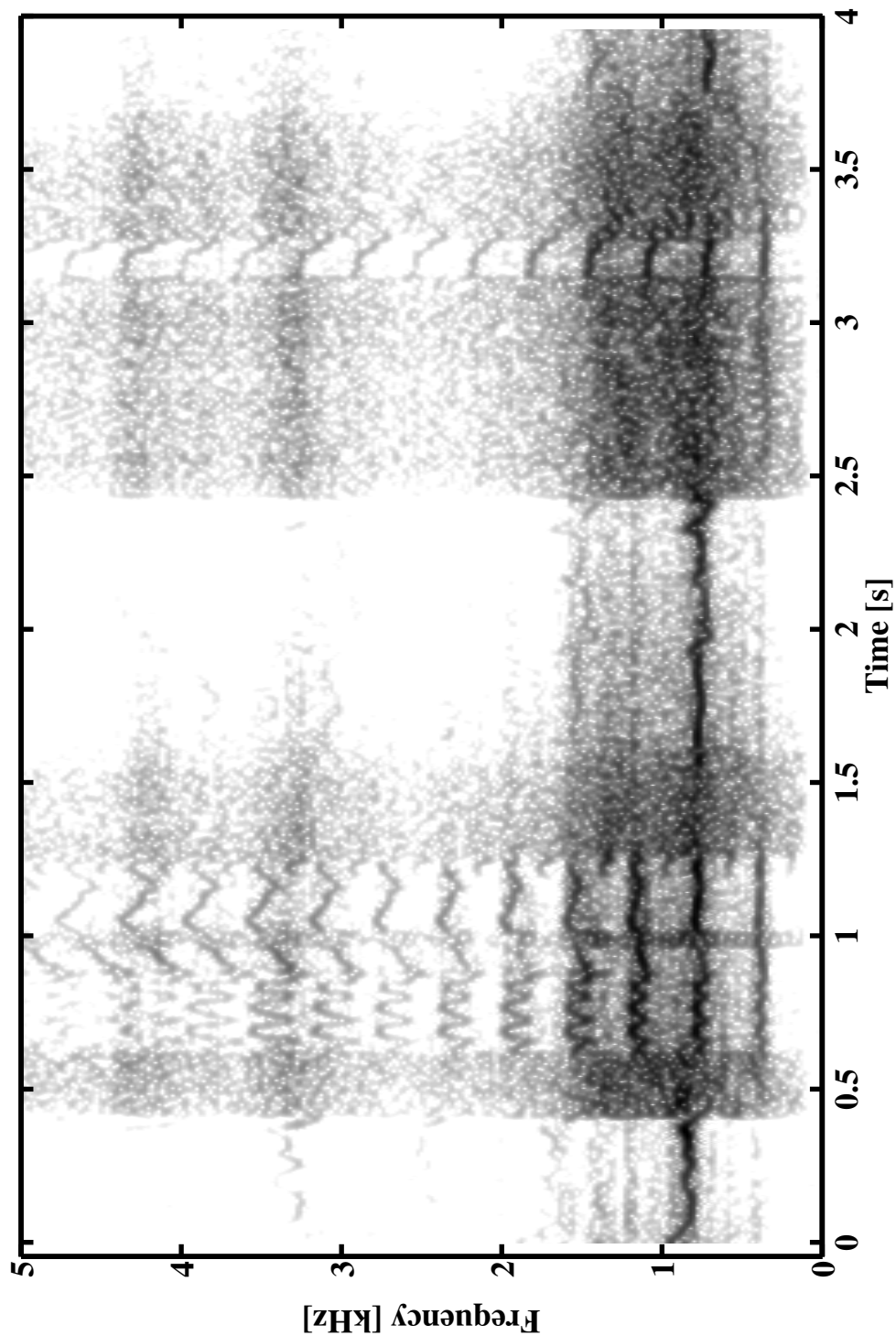


Figure 2.11: Spectrogram of sustained phonation of the voice of a singer demonstrating the repetitive intentional use of nonlinear phenomena: The left part (0.0 – 2.0 sec) shows transitions from limit cycle to chaos, followed by subharmonic oscillations and again chaos. The right part (2.0 – 4.0 sec) is qualitatively similar.

2. Physics of Voice Production

Fig. 2.11 illustrates repetitive and intentional use of nonlinear phenomena by trained vocal improvisors with healthy larynges. In the left half of Fig. 2.11 a sharp transition from a limit cycle behavior to chaotic oscillations occurs. Interspersed harmonic windows, period doubling bifurcation and again sudden onset of chaotic behavior can be observed. In the right half of Fig. 2.11 this sequence is repeated in a qualitatively similar way.

2.3.1 Coupled oscillators – the core of phonatory dynamics

As stated previously, the phonatory apparatus can be regarded as multiple coupled nonlinear oscillators (Titze, 1994a):

- Left and right vocal fold are two different oscillators.
- Within one single fold there are a number of modes (in terms of coherent spatial motion) contributing to the overall movement of the complex three-dimensional vocal fold. As previously explained, two major modes (vocal fold body movement and mucosal wave) are the crucial coherent spatio-temporal structures generating self-sustained vocal fold vibrations.
- Additional modes in longitudinal (anterior-posterior) direction or vertical (upper-lower) direction of the vocal folds add to the set of coupled oscillators.
- As the vocal folds interact with the driving airstream, aerodynamical vortex structures coupled to vocal fold oscillations are another example of oscillators of the phonatory system.
- It has been observed, that the ventricular folds, a similar valve structure above the vocal folds (see Fig. 2.6), are able to perform oscillations synchronized to the vibration of the vocal folds (Fuks et al., 1998).

The concept of coupled oscillators is crucial for the understanding and terminology of phonatory dynamics; follows a review of a few basic concepts. (For an extended description see the textbooks of Bergé et al. (1984); Glass and Mackey (1988); Jackson (1989); Schuster (1988); Strogatz (1994)).

The behavior of coupled oscillators has been extensively studied in the generic one-dimensional circle map. It describes the behavior of two nonlinearly coupled oscillators with strong dissipation using a Poincaré section (see Fig. 2.12). Two parameters, Ω and K , describe the map, where $\Omega = \omega_1/\omega_2$ is the frequency ratio of the two coupled oscillators, and K is the normalized coupling constant, a measure of the nonlinearity in the system.

The behavior of the circle map can be studied using two-dimensional bifurcation diagrams (see Fig. 2.13). For vanishing coupling K , two behaviors can be observed:

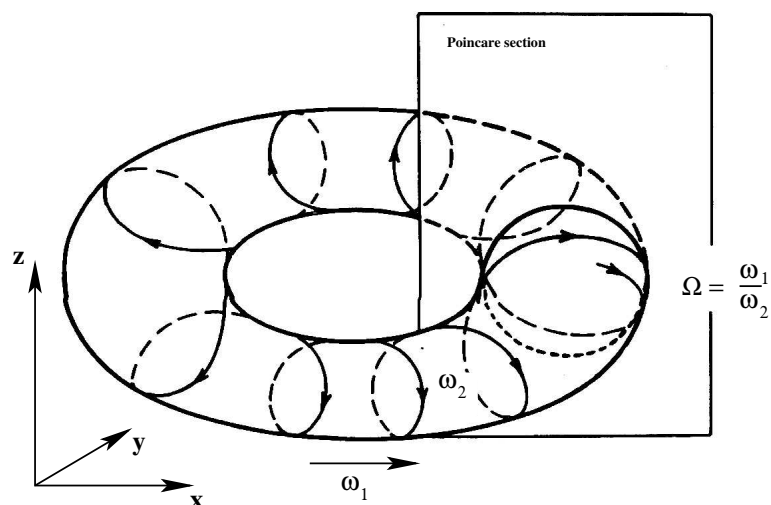


Figure 2.12: Motion on a torus in phase space. For rational frequency ratios $\Omega = \omega_1/\omega_2$ the trajectories close after a finite number of cycles (mode-locked state). For irrational frequency ratio such motion is called quasiperiodic. The trajectory never closes and asymptotically covers the whole torus, never repeating a point (from Schuster (1988)).

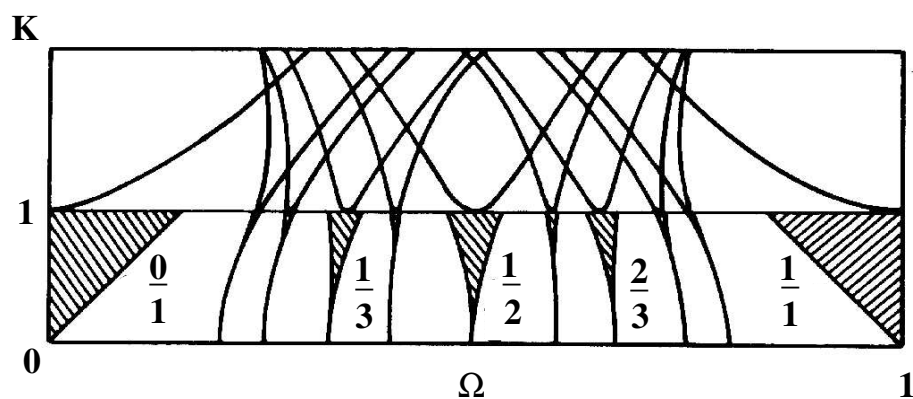


Figure 2.13: Schematic bifurcation diagram of the circle map shows mode locking within the Arnold tongues (hatched areas) for small coupling $K < 1$, merging of Arnold tongues for critical coupling $K = 1$. Note that chaos and mode locked states coexist for coupling $K > 1$. The lines (in the region $K > 1$) indicate mode locked states associated with mode locking regions (from Schuster (1988)).

2. Physics of Voice Production

The first set is characterized by rational frequency ratios $\Omega = p/q$ with $p, q \in \mathbb{N}$, where the map asymptotically produces a finite number of states. Thus, the underlying attractor of two coupled oscillators reduces to a folded limit cycle, a closed curve on the torus (Fig. 2.12). The second set is characterized by irrational frequency ratios $\Omega \in \mathbb{R}^+ \setminus \mathbb{N}$ (incommensurable independent frequencies ω_1 and ω_2), whose circle map attractor contains an infinite number of states. Trajectories of the two coupled oscillators associated with the second set lie on a torus; they never close and asymptotically fill the entire torus.

For small nonlinear coupling $0 < K < 1$, Arnold tongues (Arnold, 1965) can be found in the K - Ω bifurcation diagram (see Fig. 2.13). The tongues are finite Ω -intervals around rational frequency ratios p/q where the system performs folded limit cycle oscillations. Termed synchronization, entrainment, or mode-locking, this behavior occurs due to coupling of slightly independent oscillations. Between the Arnold tongues, there are parameter values associated with toroidal oscillations. The probability of finding tori is positive (nonzero). For fixed coupling K , the hierarchy of Arnold tongue widths is determined by the Farey tree which orders all rational numbers in the interval $[0, 1]$ according to increasing denominators (Hardy and Wright, 1938).

For the coupling of oscillators approaching $K = 1$, the entrainment regions become broader and broader. For $K = 1$, all Arnold tongues merge. Thus, the remaining nonmode-locked Ω -intervals form a Cantor set of zero measure. For $K > 1$, coexistence of different attractors, including chaotic attractors, can occur. Coexistence implies that the path to a point in the K - Ω -space is important for the asymptotical behavior of the coupled oscillators.

2.4 Experimental data acquisition of phonatory system

In this section, experimental methods for data acquisition and analysis are briefly reviewed. For an extended introduction see textbooks on phoniatrics, e.g., Titze (1994a); Wendler et al. (1996).

Traditionally, voice research uses acoustical methods to access the state of the phonatory system. Different more or less sophisticated subjective classification schemes have been developed to describe the individual acoustical perception. For instance, Nawka et al. (1994) introduced the RBH scheme (roughness, breathiness, hoarseness) for perceptual judgement of pathological voice quality in clinical practice. Ongoing research develops more or less objective methods to derive from voice and speech-related signals. Measures range from simple perturbation measures like jitter and shimmer as statistical indexes of temporal regularity, and

2. Physics of Voice Production

spectral properties like harmonics-to-noise ratio (HNR), to sophisticated measures like the ratio of glottal excitation to turbulent noise excitation (GNE), or even non-linear measures like Lyapunov exponents or attractor dimensions (e.g., de Krom, 1993; Behrman and Baken, 1997; Behrman and Agresti, 1998; Blomgren et al., 1998; Hartl et al., 2001; Kumar and Mullick, 1996; Lee et al., 1998; Narayanan and Alwan, 1995; Giovanni et al., 1999b; Matassini et al., 2000; Michaelis et al., 1997; Riede et al., 2000b, 2001; Yumoto et al., 1982; Tokuda et al., 2002). To a certain extent these measures allow a consistent classification of voice signals, but they are restricted as simplified overall measures of the behavior of the complex biomechanical-aerodynamical phonatory system.

The acoustic signals contains both information of the primary glottal vibration and the acoustical filtering of the pressure waves in the vocal tract. Although the dynamics of the glottal source cannot be uniquely estimated from the acoustical signal radiated from the mouth, the vocal output is still a simple and robust measure for analysis of fundamental frequency (pitch), time constants of frequency modulations, and occurrences of nonlinear phenomena in the glottal source signal due to bifurcations of the biomechanical-aerodynamical vocal fold system. In this thesis, acoustical signals are visualized with spectrograms (time-frequency plots) (e.g., Fig 2.9). They are regarded as spectral bifurcation diagrams where time plays the role of a control parameter (Lauterborn and Suchla, 1984; Lauterborn, 1986; Lauterborn and Parlitz, 1988). In Chapter 3, spectrograms are used for classification of nonlinear phenomena in contemporary singing in analogy to previous work on animals (Wilden et al., 1998; Fitch et al., 2002). For the study presented in Chapter 4, spectrograms are used to preselect clinical recordings of pathological voices with biphonation for further spatio-temporal analysis. In Chapter 5, spectrograms are used for an overview of the call repertoire of primates and echolocating bats. Typical fundamental frequency ranges, frequency variation, and the occurrence of nonlinear phenomena are extracted from these spectral bifurcation diagrams.

Several experimental methods have been developed that allow a more detailed description of the internal state of the larynx: Laryngoscopy is a simple visualization techniques of the static and dynamic state of the optically accessible tissue part of the mammalian larynx. In everyday clinical practice, video-stroboscopy has been established to observe superficial morphological and geometrical changes of the vocal folds (Saadah et al., 1998; Sloan et al., 1992; von Doersten et al., 1992). As stroboscopy only provides virtual slow motion, assuming a nearly periodic behavior of the vocal folds, it is not applicable in general for arbitrarily complex spatio-temporal oscillations of the vocal folds. In recent years, digital high-speed camera systems have been developed that allow the true time resolved observation of the superior medial-lateral vocal fold oscillations with reasonable high spatial

2. Physics of Voice Production

resolution. High speed glottography (HGG) (Kiritani et al., 1993; Eysholdt et al., 1994; Sundberg, 1995; Tigges et al., 1996; van As et al., 1999; Mergell et al., 1998; Berry et al., 2001; Fitch et al., 2001; Döllinger et al., 2002) will be described in more detail in Chapter 4, where complex spatio-temporal high-speed data will be analyzed. In Chapter 5 first *in vivo* high speed video recordings from neurally stimulated phonating primates (squirrel monkeys) will be presented to motivate the development of a new phonatory model for primates and bats.

Electroglottography (EGG) uses the electrical impedance changes between two electrodes applied to both sides of the neck outside the larynx to measure the contact area of the vocal folds. Like EGG, there are several methods that use probe signals transmitted from one side of the larynx to the other side to measure time variations of the glottis during phonation. Electrical current flow (as for EGG), ultrasonic waves and electromagnetic waves have been used as probe signals. In this thesis, EGG measurements are used to complement audio recordings used for preselection of pathological voices (Chapter 4), and to experimentally motivate the development of a model for sound generation in primates and bats (see literature review in Appendix of Chapter 5). The electromagnetic glottography signal uses high-frequency propagating electromagnetic waves (EM waves). It is not unambiguously associated with specific tissue oscillations because the EM wave sensor generates signals that are a function of the volume displacement of laryngeal oscillating tissue (e.g., vocal folds, tracheal wall) (Titze et al., 2000; Holzrichter et al., 1998).

Photoglottography (PGG) uses the light intensity measured by a photodiode when the glottis is transilluminated during phonation. The PGG signals measures the smallest glottal area between the vocal folds.

Inverse filtering (IF) estimates the glottal source signal from the measured acoustical signal radiated from the mouth. Linear filtering of the primary glottal signal according to linear source-filter theory has to be assumed (Titze, 1994a). Thus, the glottal airflow function corresponding to the minimum glottal area can be estimated.

Direct *in vivo* air pressure measurement with introduced pressure transducers provides a technique to determine subglottal pressure, formant frequencies (acoustical resonance frequencies, and damping factors of sub- and supraglottal cavities), and overall glottal flow (Cranen and Boves, 1985, 1987, 1988). Transglottic pressure, overall airflow, and exit jet particle velocity during phonation has been measured in an *in vivo* canine model of the larynx that is easier to access than the human larynx *in vivo* (Berke et al., 1991; Bielamowicz et al., 1999). Traveling wave velocity (wave mode of vocal fold vibrations) also has been measured in the *in vivo* canine laryngeal model (Nasri et al., 1994; Sloan et al., 1992, 1993). In Chapter 5, measured subglottal pressure ranges for phonating primates are integrated

2. Physics of Voice Production

into a new phonatory model for primates.

Magnetic resonance imaging (MRI) and computer tomography (CT) allows observation of the static vocal folds, the cartilagenous framework (Castelijns et al., 1996; Selbie et al., 2002), the ventricular folds (Honjo et al., 1985), and the vocal tract configuration changing slowly over time during articulation. (Baer et al., 1991; Dang and Honda, 1996, 1997; Story et al., 2001, 1998; Fitch and Giedd, 1999; Mathiak et al., 2000)

Apart from HGG, all visualization techniques share the problem that no true time resolved and spatially separated measurements are possible that distinguish left and right vocal fold contributions to the overall dynamics of the voice production apparatus.

Detailed muscular activity can be measured using electromyography (EMG) with thin hooked wire electrodes that are inserted directly into the intrinsic muscles of the vocal folds. During phonation, myoelastic tissue properties can be estimated measuring muscular action potentials, and neural phonation control mechanisms can be observed (e.g., Whalen et al., 1999). Data from EMG measurements in primates and bats are used for the modeling study of nonhuman phonation in Chapter 5.

In vivo examinations are complemented by *in vitro* experiments on excised larynges from humans, dogs, and primates (Brown et al., 2003; Austin and Titze, 1997; Cooper and Titze, 1985; Jiang et al., 1994, 2000, 2001a; Švec et al., 1999; Fee, 2002; Giovanni et al., 1999a; Kusuyama et al., 2001; Ouaknine et al., 2003; Yanagi and McCaffrey, 1992; Yumoto and Kadota, 1998; Numata, 1985). Easier to manipulate and to access for experimental data acquisition, excised larynx preparations are widely used to study systematically the influences of, e.g., glottal rest area, asymmetry of vocal fold tensions, and subglottal pressure on spatio-temporal behavior of vocal folds. Results from excised primate larynx experiments by Brown et al. (2003) stimulate the theoretical study in Chapter 5. Hemi-larynx experiments have been conducted to gain more insight into airflow patterns, transglottal pressure distribution, and spatio-temporal modes of vibrations of vocal folds (Alipour and Scherer, 2001; Berry et al., 2001), complementing the study presented in Chapter 4. Experimental bifurcation diagrams of excised larynges have been obtained exploring different phonatory behaviors in subspaces of the high dimensional parameter space of the phonation apparatus (Berry et al., 1996).

In *in vivo* measurements, systematic and detailed studies of airflow patterns interacting with specific glottal geometry are difficult to conduct. Replica experiments with simplified physical models of the vocal folds and a driving fluid have been performed to gain insight into phonation onset, airflow behavior and intra-glottal pressure distribution (Chan et al., 1997; Scherer et al., 2001, 1983; Vilain

2. Physics of Voice Production

et al., 2003; Shinwari et al., 2003; Titze et al., 1995; Pelorson et al., 1994). Results allow a more accurate theoretical description of the airflow and the interaction with the vocal folds.

As the phonation apparatus can be externally excited like an oscillator, resonance experiments have been performed to reveal the *in vivo* resonance structure in human larynges. Mechanical pulse-like excitation (Kaneko et al., 1972; Isshiki, 1977; Kaneko et al., 1981, 1983, 1986; Masuda, 1986) showed damped vocal fold oscillations. Using sweeps of sinusoidal mechanical external driving, resonance frequencies and damping factors together with the corresponding modes of vibration of the larynx (vocal folds, ventricular folds, aryepiglottic folds, arytenoid cartilages) (Švec et al., 2000) have been determined characterizing the mechanical spatial visco-elastic properties of the vocal folds. In Chapter 4 a numerical method for characterizing modes of vibration of the vocal folds is shown that allows a quantitative analysis of complex spatio-temporal patterns as found in resonance experiments.

From a material science point of view, tissue from different vocal fold layers has been used for rheometrical measurements. Visco-elastic tissue properties in terms of stress-strain relationships have been investigated extensively *in vitro* by, e.g., Alipour-Haghighi et al. (1989); Alipour-Haghighi and Titze (1991); Gray et al. (2000); Haji et al. (1992a,b); Min et al. (1995); Perlman et al. (1984); Titze and Talkin (1979); Titze (1976, 1993, 1994b); Chan and Titze (1999, 2000); Chan (2001); Chan and Tayama (2002); Thibeault et al. (2002). For instance, Titze (1994a) shows stress-strain relationships for vocal fold cover and muscle tissue for cyclic stretch-release experiments and estimates the active stress-strain relationship in the thyroidarythenoid muscle (vocalis muscle).

It is important to note that all stress-strain relationships are significantly non-linear, and that during cyclic stretch-release experiments the stress-strain curve follows different paths. This hysteresis behavior is related to inelastic dissipative processes in the muscular and connective tissue fibers. Visco-elastic shear properties of human vocal fold mucosa have been measured rheometrically for frequencies below phonatory frequencies (Chan and Titze, 1999, 2000; Chan, 2001). Using visco-elastic theories known from polymeric material science, extrapolations of elastic shear modulus and dynamic viscosities could be made for the phonatory frequency range. Two results are important to mention: First, the shear modulus corresponding to the stiffness of the tissue increases with frequency. Second, a shear-thinning effect has been observed, i.e. the dynamic viscosity decreases with frequency, and the dissipated energy due to heat generation in the tissue decreases with frequency, respectively. Therefore, the damping ratio, expressing the amount of damping with respect to the critical damping, is basically a flat function of frequency. Human vocal fold tissue shares this property with other biological soft

tissues (Fung, 1993). For modeling of mammalian phonation, damped oscillations of vocal fold tissue can be assumed for all frequencies.

2.5 Modeling voice production in humans and nonhuman mammals

For the last few decades normal and pathological voice production in humans has been extensively studied by more or less simple models. The complexity of the models range from low order one- or two-mass models to spatially complex Finite Element Method models. The modeling goals vary accordingly to the model type. Simple models illustrate the principle dynamical behavior of normal and pathological voices whereas complex models were developed for a complete fluid-dynamical, anatomical, and physiological description for, e.g., success prediction of phonosurgical treatment.

For nonhuman phonation, however, there are only a few studies touching the challenging topic of understanding nonhuman voice production. Questions on biological and evolutionary constraints in nonhuman phonation, metabolic costs of phonation, the role of tissue properties and of simple neurally controlled parameter variations remain difficult to be answered *in vivo* and *in vitro*. Here, *in silico* experiments with mathematical models help to bridge the gap.

Aiming at the synthesis of speech, Flanagan and Landgraf (1968) formulated a model for self-sustained vocal fold oscillations using a one-mass representation of each vocal fold acoustically coupled to a vocal tract. The aerodynamical driving forces of the bilateral mechanical mass-spring-damper system were described by the Bernoulli approximation. The airflow was assumed to be laminar, inviscid, stationary, and incompressible; hence, a potential flow with energy conservation was used. Energy transfer from the airflow to vibration of the damped one-mass oscillator was accomplished due to the inertance of the air above the glottis. Air inertance leads to a phase shift of the transglottal pressure cycle relative to the vibrating tissue. This phase shift is necessary to synchronize the driving pressure cycle with the mass vibration cycle. This mechanism required excessively high subglottal driving pressures in the one-mass model (For comparison, typical values for lung pressure, air flow rate, and other flow properties can be found, e.g., in Flanagan (1965).).

Integrating observations by Farnsworth (Farnsworth, 1940), who detected a wave-like motion of the vocal fold cover layer using analog high-speed filming of vocal fold oscillations (compare, e.g., Fig. 2.7), the next step towards a more realistic model was the formulation of a two-mass model of the vocal folds (Ishizaka and Flanagan, 1972). The wave-like shear mode of the cover layer with a phase

2. Physics of Voice Production

difference between upper and lower edge of the vocal folds improved the energy transfer from the airflow to the vibrating mass. Onset pressure for self-sustained oscillations of the model could be reduced to more realistic physiological values.

Further model generations for human and nonhuman mammalian phonation represented the biomechanical-aerodynamical phonatory system in a more or less accurate way. The motivations were diverse ranging from understanding and synthesis of normal voice production and insights of pathological phonation and embedding of phonation into nonlinear physics framework to the study of nonhuman sound production crucial for animal communication.

Advanced models with a more detailed anatomical description have been developed to study the vibratory behavior dependent on the layered inhomogeneous vocal fold tissue. Such models estimate the mechanical stress distribution in vibrating vocal fold tissue and determine effects of morphological changes of the vocal folds on spatio-temporal vibratory behavior (Titze, 1973, 1974; Titze and Talkin, 1979; Story and Titze, 1995; Jiang et al., 1998; Lous et al., 1998; Alipour-Haghighi et al., 2000).

Focusing on a realistic source signal for speech synthesis, the simple quasi-steady Bernoulli description of the driving airflow was replaced by more sophisticated methods for calculating the aerodynamical part of the phonatory system. For instance, a moving flow separation point within the glottis, based on fluid boundary layer theory, has been combined with the Bernoulli description (Pelorson et al., 1994).

The flows associated with phonation are essentially three-dimensional, turbulent, and instationary. Three-dimensional simulations are required to fully describe the complex flow (including a pulsating air jet, flow separation from the glottal walls, and turbulence), the acoustical fields, and the complex glottal geometry. The full Navier–Stokes equations have been used for the fluid mechanical description of the air flow through the glottis (Alipour-Haghighi and Titze, 1985; Liliencrantz, 1991; Alipour-Haghighi et al., 1996; de Vries et al., 2002; Zhao et al., 2002; Zhang et al., 2002a; de Oliveira Rosa et al., 2003).

First quantitative models for nonhuman sound production date back to the work of Fletcher (1988). He formulated a quantitative model for the bird syrinx using a two-mode approach of the oscillating tissue driven by Bernoulli forces. In contrast to mammals, the primary sound signal in birds is generated by the syrinx which is located at the bifurcation of the trachea into the two bronchi. Further work on the primary sound signal in birds is based either on this model (Smyth and Smith, 2002), or the two-mass model on human phonation (Fee et al., 1998; Gardner et al., 2001).

Recently, Laje et al. (2002) and Laje and Mindlin (2002) even used a slightly perturbed van der Pol oscillator as a simple generic model. This approach over-

2. Physics of Voice Production

comes the problem how to relate biomechanical tissue properties of the larynx or the syrinx, respectively, to parameters of the two-mass model. Model parameters for the van der Pol model were estimated from measured time series and spectrograms. In general, simulated time series from simple “caricature” models represent in a sense “nonlinear curve fits” of observed data of the human and animal phonatory system.

For human pathological voices, it has been shown that the dynamical complexity of simple two-mass models is high enough to invert the observed dynamics from measured time series (acoustical signals, high-speed observations) to model parameters (Mergell et al., 2000; Döllinger et al., 2002). For birds, based on observed muscular control patterns of the phonatory apparatus, Laje et al. (2002) used a generic van der Pol model for self-sustained oscillations to generate naturally sounding bird song. In general, these studies show that it is important to keep the number of model parameters small and the dynamical diversity high to succeed in fitting experimental data to phenomenological models.

As the mammalian phonatory system has been strongly conserved during evolution, the principles of human voice production can be more easily adapted to nonhuman mammalian sound production than to bird song. Using a modified two-mass model with reed-like upward extensions, Mergell et al. (1999) studied the effects of vocal membranes, thin lightweight upward extensions of the vocal folds, that can be found in a variety of echolocating bats and primates. Chapter 5 reviews this work in more detail, as it is the basis for a new dynamical model of vocal membranes in bats and primates.

The approach for a new vocal membrane model is similar to previously presented studies from literature: Acoustical data in terms of spectrograms are presented that show the variety of the sound repertoire of primates and bats. In particular, abrupt transitions between dynamically different states of the laryngeal phonatory system are shown. Additionally, anatomical data in terms of the gross morphological shape of vocal membranes in nonhuman mammals are used. The dynamical relevance of vocal membranes in the phonatory behavior of primates is shown with *in vivo* high-speed recordings. From acoustical data, only robust measures such as fundamental phonation frequencies, driving pressure ranges, and temporal dynamical patterns are used to estimate “good” parameters for the new model. Further data on muscular control patterns together with EGG measurements are integrated in the new model in terms of slow control parameter variations. The overall modeling goal is to keep the number of additional parameters small and to increase the diversity of the model as much as required in order to reproduce essential features of collected data.

2.5.1 The two-mass model as basic model

As the original two-mass model by Ishizaka and Flanagan (1972) was too complicated to be used for studies of basic dynamical features of phonation and systematic bifurcation analysis, Steinecke and Herzel (1995) proposed a simplified version that focused on the basic dynamical features of entrained oscillations of left-right and upper-lower mass oscillators. In the following, this basic model is described and discussed.

Basically, the Steinecke–Herzel model still contains the wave-like oscillatory shear mode of the vocal fold cover crucial for airstream-driven self-sustained oscillations. It neglects nonlinearities of the visco-elastic vocal fold tissues, viscous losses in the airstream, and the acoustical coupling of the glottis to the vocal tract. According to the theory of coupled oscillators described before, highly complex temporal patterns can be still observed in this simple model.

The biomechanical part of the phonatory system is described by linear coupled damped harmonic oscillators. In complex, physiologically complete, three-dimensional Finite Element Method models it has been shown that the first two basic spatial modes of vibration contribute to 98% of the total variance of all excited vibrational modes (Berry et al., 1994). As previously mentioned (see Fig. 2.7), these two modes are the in-phase oscillation of lower and upper vocal edge (vocal fold body oscillation) and the out-of-phase vibration described before as the mucosal wave mode. Choosing a mechanical model with two mechanical degrees of freedom, these two modes can be imitated with oscillations of the center of gravity of both masses and vibrations of both masses relative to each other.

The description of the aerodynamical driving force contains nonlinearities due to Bernoulli’s law for laminar incompressible stationary and inviscous potential flow. Crucial for energy transfer from the airflow to tissue vibrations is an asymmetry of the driving forces between opening and closing phase within the glottal cycle that results in a velocity-dependent force. Thus, the work $\int \vec{F} \cdot d\vec{s} = \int_0^T \vec{F} \cdot \vec{v} dt$ applied to the vocal folds by aerodynamical forces \vec{F} over the glottal cycle T does not vanish, as the force is different between opening and closing phase and thus depends on the tissue movement $d\vec{s}$ or \vec{v} , respectively.

As discussed, a negative Bernoulli pressure (relative to atmospheric pressure) adds to the elastic tissue recoil to close the glottis (Titze et al., 1994). As the Bernoulli pressure is not sensitive to the direction of the tissue vibration, this negative pressure is not sufficient to obtain an energy transfer from fluid flow to tissue vibration. The sufficient condition of breaking pressure-cycle symmetry between opening and closing is achieved in two ways: One mechanism is based on air inertance of the supraglottal vocal tract air column, as described above.

Another mechanism relies on the mucosal wave mode. That facilitates the detachment of the driving airstream from the vocal folds due to their divergent

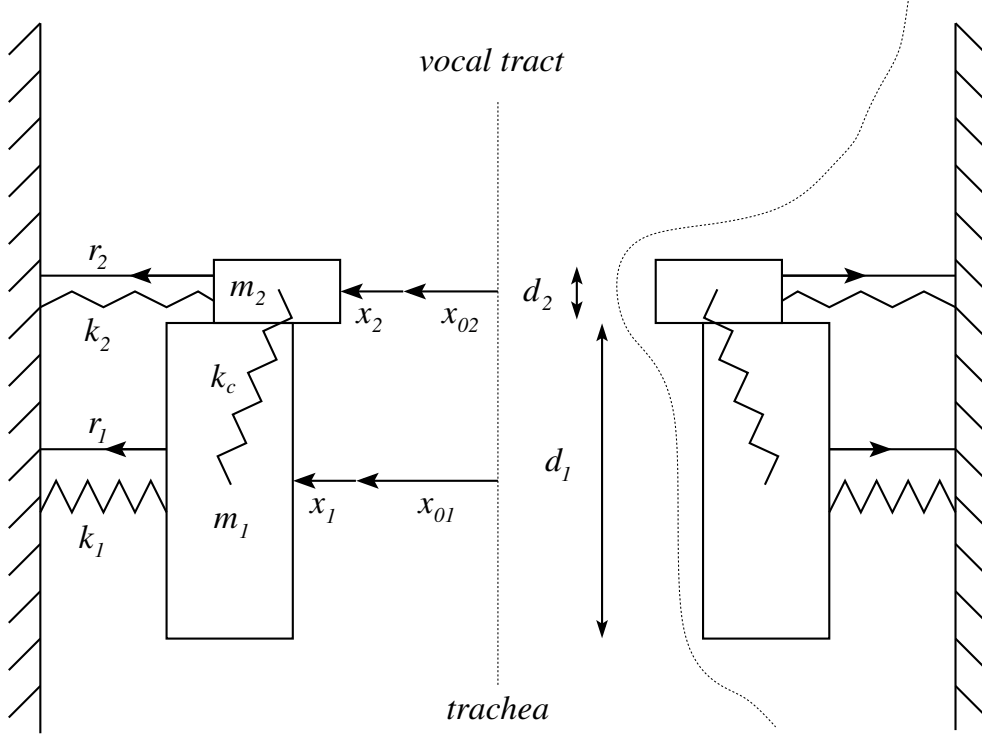


Figure 2.14: Sketch of the simplified two-mass model after Steinecke and Herzel (1995): lower and upper masses m_1 and m_2 , viscous damping r_1 and r_2 , elastic recoil k_1 and k_2 , coupling stiffness k_c , thickness of upper and lower masses d_1 and d_2 , rest position of lower and upper masses x_{01} and x_{02} , and displacement of masses from their resting position $x_1(t)$ and $x_2(t)$

shape (divergent with respect to lower-upper direction, see Fig. 2.7) during closing. In a divergent glottis, an air jet forms at a certain point along the lower-upper direction whereas for a convergent glottal shape, the flow keeps attached to the tissue surface. The simplified two-mass model incorporating the wave-like shear mode of the cover layer uses the latter mechanism, i.e. jet separation at the narrowest point of the model glottis.

Using the sketch of the biomechanical part presented in Fig. 2.14, the myoelastic part of the dynamical system can be written following basic Newton's mechanics:

$$\begin{aligned} m_1 \ddot{x}_1 + r_1 \dot{x}_1 + k_1 x_1 + k_c (x_1 - x_2) &= F_1^{\text{Bernoulli}}(t) + F_1^{\text{collision}}(t) \\ m_2 \ddot{x}_2 + r_2 \dot{x}_2 + k_2 x_2 + k_c (x_2 - x_1) &= F_2^{\text{Bernoulli}}(t) + F_2^{\text{collision}}(t) \end{aligned} \quad (2.3)$$

For the sake of simplicity, equations for only one vocal fold side are given as the basic model is symmetrical. The dynamics $x_1(t)$ and $x_2(t)$ of the two masses

2. Physics of Voice Production

m_1	m_2	k_1	k_2	k_c	r_1	r_2	l
0.125	0.025	0.08	0.008	0.025	0.02	0.02	1.4
c_1	c_2	a_{01}	a_{02}	d_1	d_2	ρ	p_s
$3k_1$	$3k_2$	0.05	0.05	0.25	0.05	0.00113	0.008

Table 2.1: Set of standard parameters for the simplified symmetric two-mass model simulating normal male phonation (in the unit system cm, ms, g) (Steinecke and Herzel, 1995)

is governed by equations of two coupled damped harmonic oscillators driven by nonlinear forces. The driving forces of airflow through the model $F_i^{Bernoulli}$ and the collision of left and right vocal folds $F_i^{collision}$ contain the only nonlinearities.

The Bernoulli forces are proportional to the dynamical pressures p_i acting on the surface areas ld_i of the model masses:

$$F_i^{Bernoulli}(t) = l d_i p_i(t) \quad (i = 1, 2) \quad (2.4)$$

The collision forces act as additional restoring force (additional spring constants c_i) during vocal fold contact which corresponds to an overlap of left and right model masses. If the masses penetrate, the glottal areas a_1 and a_2 are negative and are proportional to the penetration depths $|a_i(t)|/2l$:

$$F_i^{collision}(t) = c_i \frac{|a_i(t)|}{2l} H(-a_i(t)) \quad (i = 1, 2) \quad (2.5)$$

where the Heaviside function $H(a)$ is defined as:

$$H(a) = \begin{cases} 0 & , a < 0 \\ 1 & , a \geq 0 \end{cases} \quad (2.6)$$

To specify the simplified geometry, glottal geometry parameters have to be defined (see Fig. 2.14): l is the vocal fold length, d_i are the vertical depths of the model masses, and $a_{0i}^{left} = l x_{0i}$ the glottal rest areas related to the rest positions x_{0i} of the masses.

The instantaneous glottal areas are $a_i^{left}(t) = a_{0i}^{left} + l x_i(t)$. With the corresponding displacements x_i^{left} and x_i^{right} for left and right vocal fold model sides the total cross-sectional areas are $a_i(t) = a_i^{left}(t) + a_i^{right}(t)$.

Using the unit system {cm, ms, g} all standard parameters for model geometry and visco-elastic properties are shown in Table 2.1 . Traditionally, in voice research, the pressure unit $[p_s] = \text{cmH}_2\text{O}$ is used. This is converted as:

$$1 \text{ cmH}_2\text{O} = 9.81 \cdot 10^{-4} \frac{g}{cm \cdot ms^2} \approx 10^{-3} \frac{g}{cm \cdot ms^2} \quad , \quad (2.7)$$

2. Physics of Voice Production

with the density of water $\rho_{H_2O} = 1.0 \text{ kg/l}$.

The pressure driving this model is derived from Bernoulli's law describing potential flow (laminar, incompressible, stationary flow without viscous losses). The subglottal pressure p_s produced by the lungs is transformed into air pressure p_i and flow velocities U/a_i ($i = 1, 2$) within the glottis corresponding to the cross-sectional area function (here a step function with values a_1 and a_2) within the glottis. The volume flow U is constant within the glottis due to incompressibility. For a small glottis orifice-to-tube area ratio, the volume flow in the subglottal tube is negligible. The pressure above the glottis in the vocal tract p_T is set to the surrounding atmospheric pressure $p_T = 0$ and therefore neglects acoustical pressure waves in the vocal tract. Similarly, acoustical pressure waves in the subglottal tube are neglected. This pressure p_s describes the transglottal pressure drop, the static pressure gradient across the glottal orifice. ρ describes the uniform air density at body temperature and corresponding humidity.

Thus, for an open glottis, Bernoulli's law states:

$$p_s = p_1(t) + \frac{1}{2}\rho \left(\frac{U(t)}{a_1(t)} \right)^2 = p_2(t) + \frac{1}{2}\rho \left(\frac{U(t)}{a_2(t)} \right)^2 \quad (2.8)$$

An air jet is assumed to form at the narrowest point $a_{min} = \min(a_1, a_2)$ of the glottis. Thus, subglottal pressure can be related to the total volume flow via:

$$p_s = \frac{1}{2}\rho \left(\frac{U(t)}{a_{min}(t)} \right)^2 \quad (2.9)$$

The jet is assumed to have the cross-sectional area a_{min} .

Then, the constant glottal volume flow $U(t)$ can be written as:

$$U(t) = \sqrt{\frac{2p_s}{\rho}} a_{min}(t) H(a_{min}) \quad (2.10)$$

Assuming jet separation, it follows that $p_2(t) = p_T = 0$ and:

$$p_1(t) = p_s \left[1 - \left(\frac{a_{min}(t)}{a_1(t)} \right)^2 H(a_{min}) \right] H(a_1) \quad (2.11)$$

Note, that the Heaviside functions are introduced to fulfill the open glottis condition.

The two-mass model equations are integrated with a numerical 4th order Runge-Kutta scheme. The step size has been fixed to a sampling rate of $f_s = 20000 \frac{1}{s} = 20 \frac{1}{ms}$. An integrator scheme with adaptive step size fails due to discontinuities of the derivative of the driving forces during contact of left and right masses.

2. Physics of Voice Production

The standard initial conditions for the numerical integration are:

$$\begin{aligned} x_{1l}(t_0) &= x_{1r}(t_0) = 0.1 \\ x_{2l}(t_0) &= x_{2r}(t_0) = 0.0 \\ \dot{x}_{1l}(t_0) &= \dot{x}_{1r}(t_0) = 0.1 \\ \dot{x}_{2l}(t_0) &= \dot{x}_{2r}(t_0) = 0.0 \end{aligned} \tag{2.12}$$

with the subscripts “l” and “r” indicating left and right vocal fold mass displacements.

For the numerical evaluation of the Heaviside function $H(a)$, the approximation

$$H(a) := \begin{cases} 0 & , a < 0 \\ \tanh(50 a) & , a \geq 0 \end{cases} \tag{2.13}$$

is used.

The two-mass model described above is only valid for left-right symmetrical vocal folds. In order to study effects of laryngeal asymmetries frequently observed in pathological voices (Sloan et al., 1992; Woodson, 1993; Wittenberg et al., 2000) and in normal healthy voices (Ward et al., 1969; Herzel and Reuter, 1997; Mergell and Herzel, 1997a), asymmetry factors have been introduced. Motivated by the theory of coupled oscillators with different vibration frequencies, asymmetry factors detune left and right vocal fold masses and stiffnesses, e.g., $Q = m_{il}/m_{ir} = k_{ir}/k_{il} = c_{ir}/c_{il}$. Physiologically, the asymmetry factors can be interpreted as laryngeal asymmetries of vibrating masses, differences in tissue elasticities and viscosities, and vocal fold geometry.

Systematically detuning the eigenfrequencies of left and right vocal fold oscillators, bifurcation diagrams can be obtained revealing different oscillatory regimes with subharmonic oscillations, tori, and deterministic chaos. The coupling between left and right vocal fold oscillators can be adjusted by varying subglottal pressure and glottal rest areas. Temporally complex phonatory behaviors in pathological and normal healthy voices can be qualitatively understood by time-dependent detuning of left and right vocal fold oscillators (vocal irregularities in pathological voices have been observed, e.g., by Smith et al. (1992) and Steinecke and Herzel (1995)).

According to the theory of coupled oscillators, a variety of nonlinear phenomena for different parameter combinations have been observed in this model. Thus, high-speed observations from normal and pathological voices could be reproduced with this simplified two-mass model by adjusting asymmetry factors detuning left and right eigenfrequencies, subglottal pressure, and glottal rest areas (Mergell et al., 2000; Döllinger et al., 2002).

It is worth noting that even the symmetrical two-mass model is capable of deterministic chaotic behavior if high subglottal pressure values are used to drive the model (Jiang et al., 2001b; Jiang and Zhang, 2002b).

2. Physics of Voice Production

Several extensions have been made to model the influence of further oscillators on the vocal fold dynamics. Using a single tube approximation, the effect of vocal tract resonances on the disentrainment of left and right vocal folds has been studied (Mergell and Herzel, 1997b). In another study, Lous et al. (1998) coupled a model with a smooth area function, otherwise similar to the two-mass model, to a simple representation of the vocal tract and the subglottal trachea.

So far, model extensions covered either the acoustical coupling to sub- and supraglottal cavity resonances, the variability of vocal fold oscillations in longitudinal (anterior-posterior) direction (e.g., in terms of a 16-mass model (Titze, 1973, 1974), or Finite Element Method models for a complete anatomical representation of the vocal folds (Alipour-Haghighi et al., 2000)).

A first study on the effects of upward extensions has been done by Mergell et al. (1999). This study was motivated by human voice pathologies with morphological changes (such as cysts, polyps, Reinke’s edema) and by vocal membranes (thin upward lightweight membranous extensions of the vocal fold cover found in many nonhuman mammalian species). Chapter 5 develops an extension of this model and examines the dynamical relevance of vocal membranes for nonhuman mammalian voice production.

Chapter 3

Nonlinear Phenomena in Contemporary Vocal Music

This chapter is a revised version of Neubauer, J., Edgerton, M., and Herzel, H. (2004), “Nonlinear phenomena in contemporary vocal music,” *J. Voice* In press. (Neubauer et al., 2004)

Complex and multiphonic voice signals of vocal improvisors are analyzed within the framework of nonlinear dynamics. Evidence is given that nonlinear phenomena are extensively used by performers associated with contemporary music. Narrow-band spectrograms of complex vocalizations are used to visualize the appearance of nonlinear phenomena (spectral bifurcation diagrams). Possible production mechanisms are discussed in connection with previous research, personal performance and pedagogical experience. Examples for period doubling, biphonation and irregular aperiodic phonation in vocal sonorities of contemporary vocal improvisors are given, and glottal whistle production encompassed with biphonation and triphonation is shown. Furthermore, coincidences of harmonics-formant matching associated with abrupt transitions to subharmonics and biphonation in the vocal output are provided. This also shows the recurrent use of nonlinear phenomena by performers. It is argued that mechanisms such as source-tract coupling or vocal fold desynchronization due to asymmetry are used in a reproducible way for musical tasks.

3.1 Introduction

Composers, performers and listeners of contemporary classical music have long recognized the vitality of complex multiphonic instrumental and vocal sonorities. However, the theoretical understanding of these complex sounds relied mainly upon the methods of mechanical reproduction (i.e. fingering charts with embouchure indications), while scientific questions were largely avoided (except, e.g., Gibiat and Castellengo, 2000). Beginning in the early 1980's theories of nonlinearity were applied to complex musical signals. This led some composers and musicians to reconceptualize their understanding of the elements involved in the production of sound.

In this chapter, complex and multiphonic voice signals from vocal improvisors will be analyzed within the framework of nonlinear dynamics. Voice samples from solo vocal improvisations are classified as being harmonic voice, subharmonics, biphonation (two independent pitched melodies) or irregular aperiodic behavior.

Nonlinear phenomena in voice are widely observed in newborn cries (Mende et al., 1990; Robb and Saxman, 1988), pathological voices (Nonomura et al., 1996; Terrio and Schreibweiss-Merin, 1993; McKinney, 1982; Herzel et al., 1994; Marasovich et al., 1993), animal vocalizations (Wilden et al., 1998; Fee et al., 1998; Brown and Cannito, 1995; Fletcher and Tarnopolsky, 1999), and occasionally in speech (Klatt and Klatt, 1990; Dolansky and Tjernlund, 1968). Further, it has been reported that nonlinear vocal phenomena carry functional and communicative relevance for animals and humans (Kohler, 1996; Fitch et al., 2002).

In a previous case study, Paul Ward (Ward et al., 1969) investigated vocalizations of a teenage subject who had the ability to produce biphonation. This subject achieved such behaviors through asymmetrical control of the left and right vocal folds. Captured on high speed photography and cinefluorography, the subject demonstrated the capacity to produce two different, but simultaneous pitch movements. Otherwise the subject had a completely normal voice. Further, she had the proficiency to produce such behaviors within clearly identifiable musical scales and not simply as contour relationships.

This study is representative of a growing body of evidence that suggests normal larynges have the ability to produce nonlinear phenomena including subharmonics (Large and Murry, 1978; Kaufman, 1975; Smith et al., 1967; Mazo et al., 1995), biphonation (Gerratt et al., 1987, 1984; Ward et al., 1969; Tigges et al., 1997) and aperiodic, irregular behavior (deterministic chaos) (Mazo et al., 1995; Kavasch, 1980; Barnett, 1972; Lee et al., 1998). In addition, it has been documented in the musical literature that these and many other nonstandard vocal outputs are widely used by performers of contemporary music (Large and Murry, 1979; Anhalt, 1984; Wishart, 1983; Jensen, 1979; Clark, 1985; Newell, 1970; Barnett, 1972; Chase, 1975; Edgerton et al., 1999; Gottwald, 1998; Schnebel, 1972; Liska-Aurbacher,

3. Nonlinear Phenomena in Contemporary Vocal Music

2000; Chadabe, 1997).

Here a framework is provided to characterize these frequently used nonstandard vocalizations in contemporary vocal music and discuss possible physiological mechanisms. The extensive use of temporally and spectrally complex vocalizations by performers of contemporary vocal music is documented within the framework of nonlinear dynamics. Musically, it is argued that nonlinear phenomena in the voice can be consciously used for artistic purposes.

3.1.1 Nonlinear phenomena

The human voice production system can be considered as multiple nonlinear coupled oscillators. For instance, oscillators can be left and right vocal folds, different modes of vibrations within single folds, the ventricular folds, the epiglottis, or aerodynamical oscillators (Titze, 1994a). The vocal folds disrupt the outward flowing airstream using a wave-like vibratory motion. This oscillation is sustained by the airflow provided by alveolar (lung) pressure together with visco-elastic forces within the larynx. The theoretical description of normal periodic phonation requires already a nonlinear theory, as, e.g., phonation onset from prephonatory standstill is described as an instability of the rest state resulting in self-sustained oscillations. Furthermore, nonlinear laws govern the dependence of normal phonation upon the driving pressure, rate of airflow, amplitude of vocal fold oscillation, stress-strain properties of vocal fold tissue and vocal fold collision.

This study focuses on vocal behaviors of vocal improvisors that deviate from normal periodic phonation in healthy larynges. In the following discussion the term nonlinear phenomena will consequently refer to subharmonics, biphonation, and irregular aperiodic phonation (associated with deterministic chaotic behavior).

Basic principles

Understanding the dynamics of nonlinearly coupled oscillators requires an introduction to a few basic concepts of nonlinear dynamics. A detailed introduction to nonlinear dynamics can be found e.g. in Bergé et al. (1984); Glass and Mackey (1988); Herzl et al. (1994); Kaplan and Glass (1995). Analysis of real-world phenomena using methods from nonlinear dynamics is based on the phase space to describe the state and behavior of a system. The phase space is built from the dynamical variables necessary to determine the state of a (nonlinear) system. At every moment, the behavior of a system may be represented by a single phase space point.

After initial transients systems reach a particular dynamic regime. This regime corresponds to a certain geometrical object in phase space termed an attractor. Four types of attractors have been identified: 1) steady state, a behavior whose

3. Nonlinear Phenomena in Contemporary Vocal Music

dynamical variables are constant; 2) limit cycle, a periodic behavior (repeating itself continuously); 3) torus, a two-dimensional object in phase space that results from the superposition of two independent oscillations ; 4) chaotic attractor, a aperiodic behavior that never repeats but stays within a limited space – in phase space typically a fractal geometrical object.

Attractors govern the dynamics for constant external parameters such as vocal fold tension or subglottal pressure. Often these parameters vary slowly compared to the typical dynamical behavior such as vocal fold vibrations. A slowly varying parameter can induce a sudden transition between different attractor types, a behavior called bifurcation. Of particular relevance are: 1) Hopf bifurcation, a transition from a steady state to a limit cycle; 2) period doubling bifurcation, a transition from a limit cycle to folded limit cycle; 3) secondary Hopf bifurcation, a transition from a limit cycle to a torus, due to the excitation of another independent oscillation; 4) onset of deterministic chaos, such that a small parameter shift induces a jump to aperiodic oscillation. Often cascades of subharmonic bifurcations and tori are precursors of deterministic chaos.

Applications to voice production

The framework of nonlinear dynamics can be used to classify complex vocal behavior of (contemporary) vocal improvisors. Steady state behavior occurs when the vocal folds are at rest. Then as subglottal air pressure increases, a Hopf bifurcation changes the steady state attractor into a limit cycle as the vocal folds begin to vibrate in a normal, periodic way. Often during speech and song, period doubling bifurcations occur which can be seen as subharmonic oscillations. Subharmonics may be classified as a special type of limit cycle which appears via transition from a periodic oscillation to an oscillation with alternating amplitudes. In phase space, subharmonic oscillations are represented by a folded limit cycle. Less frequent, though still seen in speech and song, is a phenomenon featuring two or more independent frequencies. This phonation, in phase space classified as a torus, may be produced with left-right asymmetrical vocal fold vibrations and has been termed biphonation (Sirviö and Michelsson, 1976; Herzel et al., 1994). As mentioned above, subharmonics and tori often are precursors of deterministic chaos. In this case, the behavior is aperiodic, irregular and complex, but not necessarily random.

In order to provide evidence of deterministic chaos, phase space analysis is required (Herzel et al., 1994; Mende et al., 1990). In earlier studies attractor dimensions and Lyapunov exponents have been estimated from voice signals (Behrman and Baken, 1997; Behrman, 1999; Giovanni et al., 1999b; Jiang et al., 2001b; Matassini et al., 2000; Narayanan and Alwan, 1995). A comprehensive phase space analysis of different attractors of voice signals is beyond the scope of this study (for

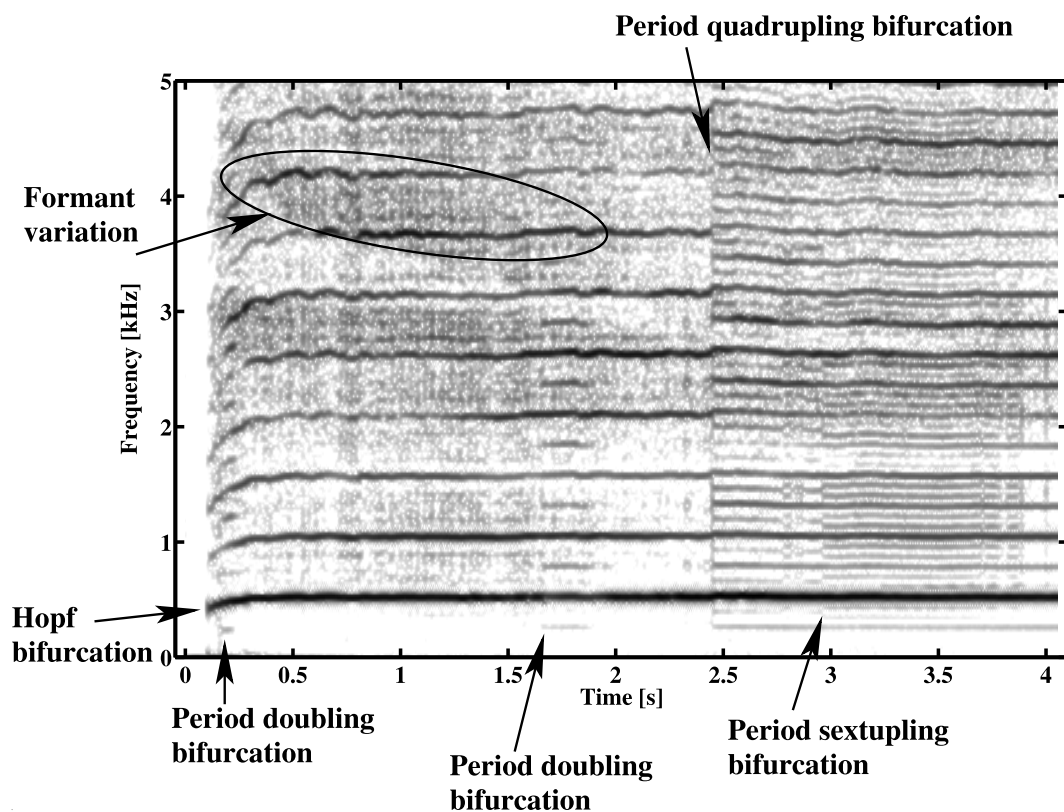


Figure 3.1: Spectral bifurcation diagram (spectrogram, sonagram) revealing a period doubling cascade from a harmonic behavior with a pitch of about 500 Hz to subharmonics at $\frac{1}{2}$, $\frac{1}{4}$ or $\frac{1}{6}$ of the pitch. The sample was taken from a recording of the solo vocal improvisation *investigazioni (diplofonie e triplofonie)* by the vocal improviser Demetrio Stratos (Stratos, 1978).

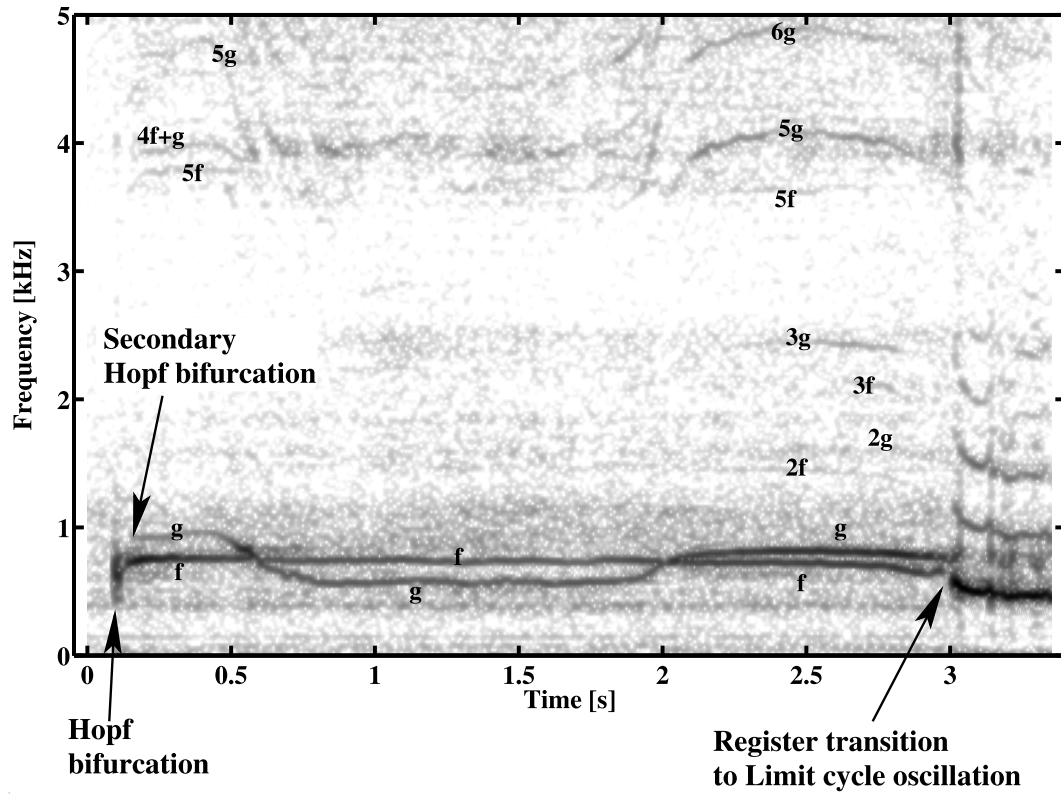


Figure 3.2: Spectral bifurcation diagram showing biphonation with two independent frequencies (two independent pitched melodies) termed f and g and various linear combinations of these frequencies. The sample is from a recording of the solo vocal improvisation *passagi 1,2* by the vocal improviser Demetrio Stratos (Stratos, 1978).

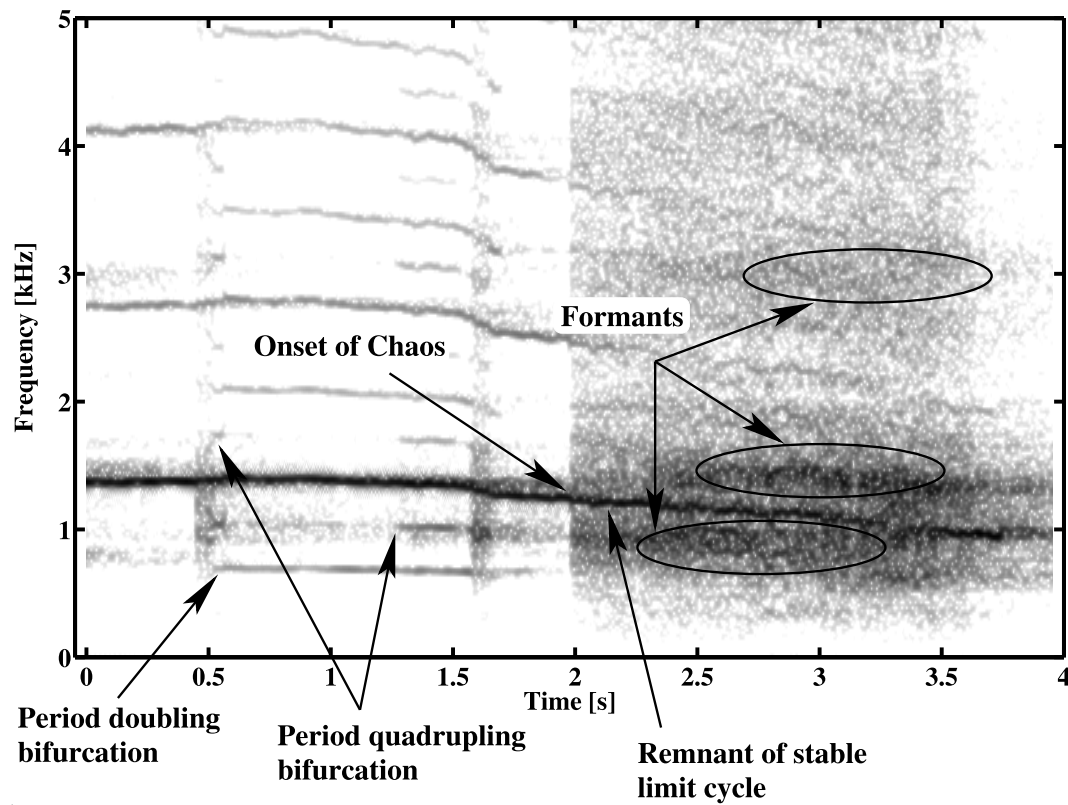


Figure 3.3: Spectral bifurcation diagram displaying the transition from harmonic behavior via period doubling, period quadrupling to an irregular, noise-like segment. Within the broadband noise-like segment the vocal tract filter is revealed by different shadings. The sample was taken from a recording of the solo vocal improvisation *entre nosotros - epitafio a las ballenas* by the vocal improviser Fatima Miranda (Miranda, 1992).

3. Nonlinear Phenomena in Contemporary Vocal Music

attractor dimensions, Lyapunov exponents see, e.g., Titze et al., 1993; Herzel et al., 1998; Kumar and Mullick, 1996; Lee et al., 1998). Fortunately, many conclusions about the dynamics can be drawn from narrow-band spectrograms. These can be interpreted as spectral bifurcation diagrams where time is considered a control parameter (cf. Lauterborn and Suchla, 1984; Lauterborn, 1986; Lauterborn and Parlitz, 1988): Subharmonics (e.g., see Fig. 3.1) appearing after period-doubling bifurcations correspond to parallel lines in between harmonics. These occur typically at multiples of $1/2$, $1/3$ or $1/4$ of the original pitch, generally at n/m of the original pitch, where n and m are small integers. In the case of biphonation (e.g., Fig. 3.2), seemingly independent spectral components with no simple ratio (such as $1/2$, $1/3$, that are modulated differently or move independently) appear in the spectrum. Finally, chaos (see Fig. 3.3) is characterized by broadband noise-like segments that appear via abrupt transitions. Distinct spectral peaks embedded in the broadband spectrum appear due to recurrent almost periodic behavior within the chaotic segment. Deterministic chaos is often interrupted by windows of normal periodic phonation or subharmonic phonation.

3.2 Musical phenomena

Nonlinear phenomena have been well-documented in many disciplines, but regularly ignored by the mainstream of performers and composers of contemporary classical vocal music who treat the voice with much the same musical aesthetic and principles of phonation that existed during the 18th and 19th centuries. So that while instruments have incorporated multiphonics, complex inharmonic signals and other transient phenomena (Gibiat and Castellengo, 2000; Maganza et al., 1986), the voice has remained for the most part the carrier of simple melodic formulæ.

Nonlinear phenomena have been identified as important instrumental resources for experimental new music since the late 1950s when Bruno Bartolozzi attempted to systematize a framework for the production of multiphonic sonorities for woodwinds (Bartolozzi, 1969). These innovations, combined with the continuing technical development and conceptual radicalism of electroacoustic music, suggested to many that the voice should share in the wealth of new sound and its construction. Therefore, during the 1960s, composers such as Dieter Schnebel, Luciano Berio, John Eaton, Giacinto Scelsi, György Ligeti, Kenneth Gaburo, Pauline Oliveros, Sylvano Bussotti, Robert Erickson and Mauricio Kagel began to explore the production and organization of non-standard vocal music (Griffiths, 1981). However, most of this work did not attempt to systematically utilize nonlinear phenomena. This is completely understandable, for unlike instruments, the human voice cannot easily be taken apart and put back together. As standardized fingering charts for vocal sound production within the larynx were lacking, most composers at-

3. Nonlinear Phenomena in Contemporary Vocal Music

tempted to explore performance technique and expression through phonetic based articulatory procedures. To a far lesser degree they tried to use combinations of multiple vocal sound sources, combining primarily harmonic with inharmonic input, or special phenomenon, such as subharmonics or overtone singing. In the last two decades of the 20th century, the active search for expression through the vocal instrument has subsided and new compositions again feature mostly traditional modes of vocal performance. However, nonlinear phenomena of the voice are extensively used by performers who are often classified as vocal improvisers.

For this study, complex vocalizations from solo vocal improvisations by vocal improvisors Greetje Bijma (Bijma, 1989), Jaap Blonk (Blonk, 1997), Anna Homler (Homler, 1989), Fatima Miranda (Miranda, 1992), David Moss (Moss, 1989), and Demetrio Stratos (Stratos, 1978) were chosen. Examples for the appearance of nonlinear phenomena in these singers are shown. As detailed measurements are difficult to perform, only ideas about possible production mechanisms are provided. This study argues that these mechanisms are used in a reproducible way for musical tasks.

3.3 Material and Methods

Narrow-band spectrograms are presented as time-frequency plots of sound samples of vocal improvisors. Narrow-band spectrograms could be interpreted as spectral bifurcation diagrams where time is considered a control parameter (Lauterborn and Suchla, 1984; Lauterborn, 1986; Lauterborn and Parlitz, 1988). The samples were recorded at a sampling rate of 44100 Hz with a resolution of 16 bits/sample. For the calculation of the spectrograms, a window size of 2048 sample points was used for the Fourier analysis. Therefore, the frequency resolution could be calculated as $\frac{44100}{2048} \text{ Hz} \approx 21.5 \text{ Hz}$, which corresponded to the smallest frequency of the Fourier analysis. A Hanning window for the Fourier analysis was used to account for finite window size effects. The sound signals were centered about their mean values. A temporal overlap of 2000 sample points was used for the sliding window Fourier analysis. If not otherwise specified, the dynamical range of the spectrograms was chosen as 60 dB sound intensity level. All spectrograms were normalized with respect to their maximum intensity level. A high spectral resolution of 21.5 Hz rather than a high temporal resolution was chosen for the calculation of the spectrograms. It was assumed that *in vivo* physiological parameters, such as vocal tract configuration or subglottal pressure, changed slowly over time (Riviere et al., 1998). This variation could still be resolved by the low temporal resolution of the spectrograms.

All samples were recordings from single singers only. The singers were vocal improvisors performing solo vocal improvisations with no sound processing. All

3. Nonlinear Phenomena in Contemporary Vocal Music

recordings consisted of the pure microphone signal without postprocessing, neither dynamically (dynamical compression or enhancement) nor spectrally (filtering, pitch shifting, chorusing, phasing). The sound samples were digitally copied from publicly available CD recordings.

In musical terms, the articulatory gestures of vocal improvisors could be described as nonturbulent open vowel-like sonorities. Musically, improvisors aimed to reach an artistically pleasing and highly cultured state.

This study was based on voice samples that were chosen from 300 different samples of 7 different singers. The shown representative subset of samples was chosen due to preliminary perceptual evaluation concentrating on episodes with independent frequencies, fast transitions, pitch instabilities, whistle-like vocalizations and broadband sounds. With respect to contemporary music, voice samples were selected that combined artistically interesting multi-tone complex vocal sonorities with dynamically interesting nonlinear phenomena.

3.4 Results

In Fig. 3.1 an example of period doubling with windows of inharmonicity is shown. At 1.6 s a period doubling bifurcation (P2) occurred indicated by an additional stack of harmonics interspersed between the existing harmonics. At about 2.5 s a period quadrupling bifurcation (P4) occurred, at 2.7 s a bifurcation to a period sextupling (P6) appeared – a period six times the fundamental pitch. After a P4 bifurcation (2.8 s) the system jumped via a P6 bifurcation again (3.0 s). The fundamental frequency of about 500 Hz suggested that the vocal folds of the male singer vibrated in the falsetto register of a male voice. This finding was consistent with the perceptual judgement.

In Fig. 3.2 an example of biphonation is shown, the occurrence of two independent frequency contours that may be perceived as two different pitch sequences (melodies). At about 0.6 s and about 2.0 s the time course of the two frequencies crossed, which supported the observation of independence of the two frequencies. In addition to the two independent frequency contours f and g , frequency components could be observed which could be explained by linear combinations of f and g . During the biphonic episode the amplitudes of the combination frequencies were relatively small. This could indicate that either the amount of coupling (i.e. the nonlinearities responsible for combination frequencies) between the oscillators was low. Alternatively, the vocal tract filter could have reduced the intensity of combination frequencies. A third interpretation was, that the harmonics of the two frequencies f and g were low. It was carefully checked, that this observation was not an effect of the reduction of the displayed dynamical range of the spectrograms. At about 3.0 s a register transition to a nonbiphonic falsetto voice occurred.

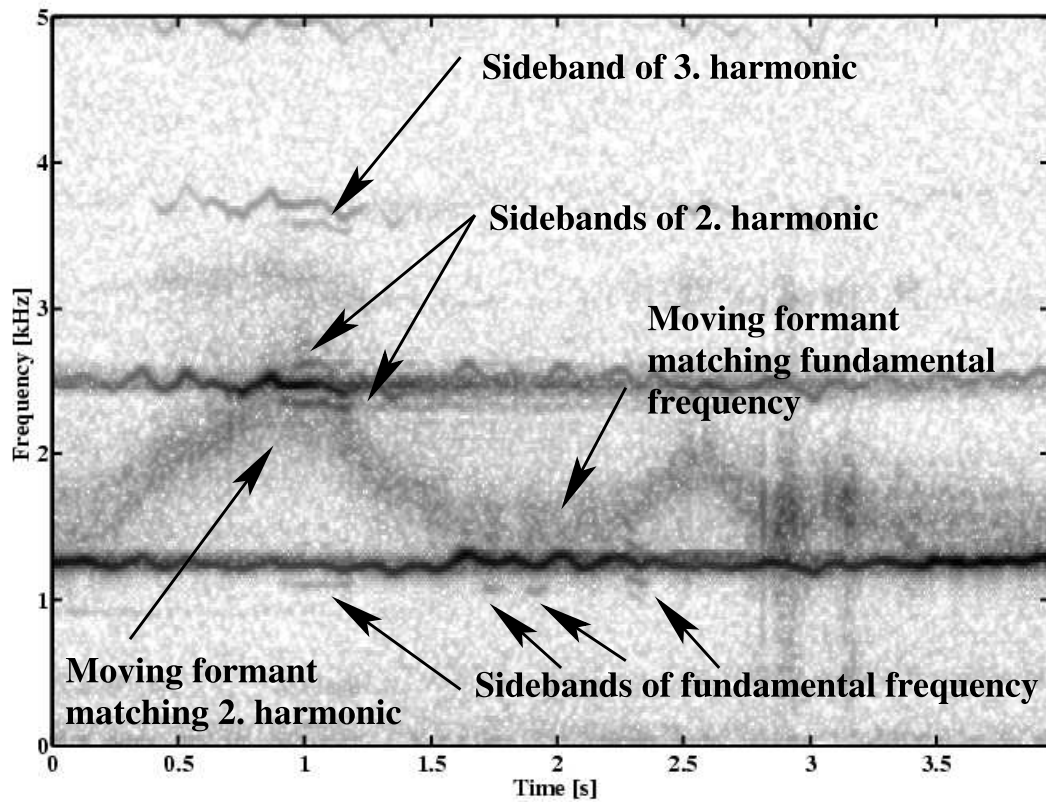


Figure 3.4: Spectral bifurcation diagram showing coincidence of formant matching harmonic associated with sideband modulations of harmonic component. Here, the dynamical range for the gray scale coding of power spectrum intensities was chosen as 90 dB sound intensity level. The sample was taken from a recording of the solo vocal improvisation *in principio* by the vocal improviser Fatima Miranda (Miranda, 1992).

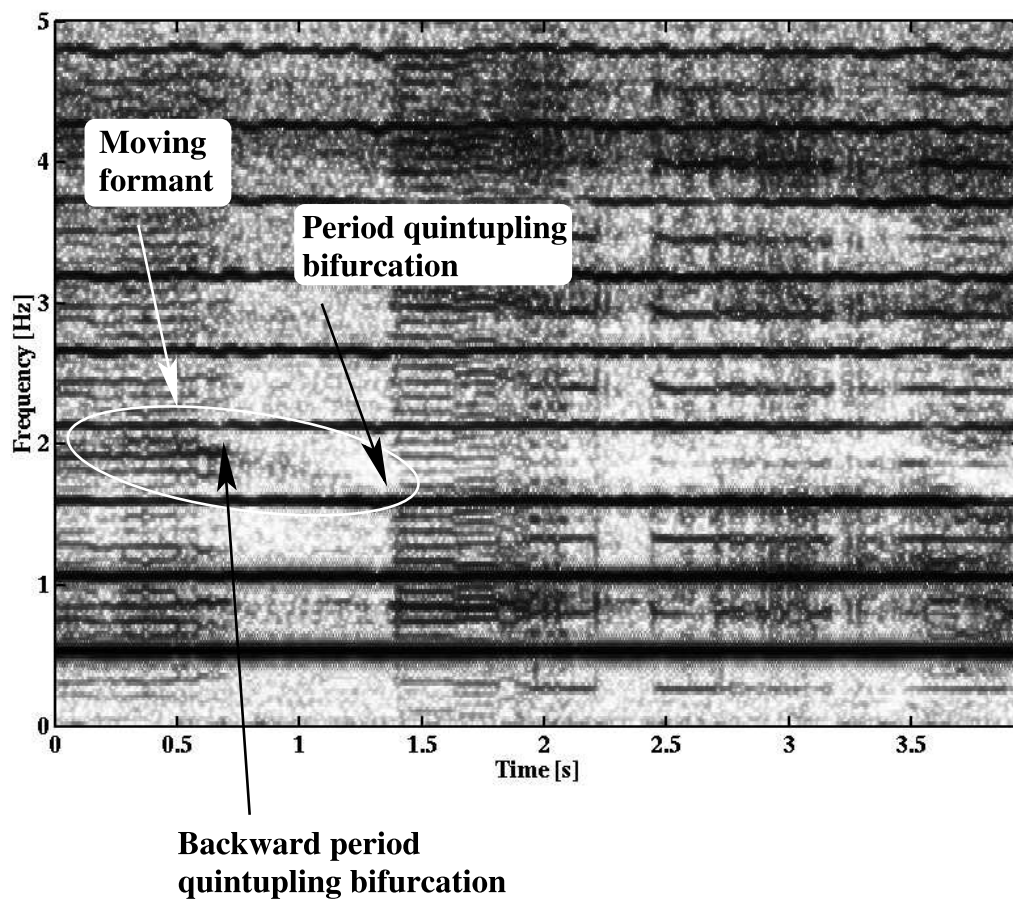


Figure 3.5: Spectral bifurcation diagram revealing recurrent source-formant interaction: Formant-induced P5 bifurcation with five subharmonics vanishes and reappears when a formant frequency matches the fourth or the third harmonic, respectively. The dynamical range for the gray scale coding of power spectrum intensities was chosen as 90 dB sound intensity level. The figure contrast and intensity was manually adjusted to emphasize formants. The sample was taken from a recording of the solo vocal improvisation *investigazioni (diplofonie e triplofonie)* by the vocal improviser Demetrio Stratos (Stratos, 1978).

3. Nonlinear Phenomena in Contemporary Vocal Music

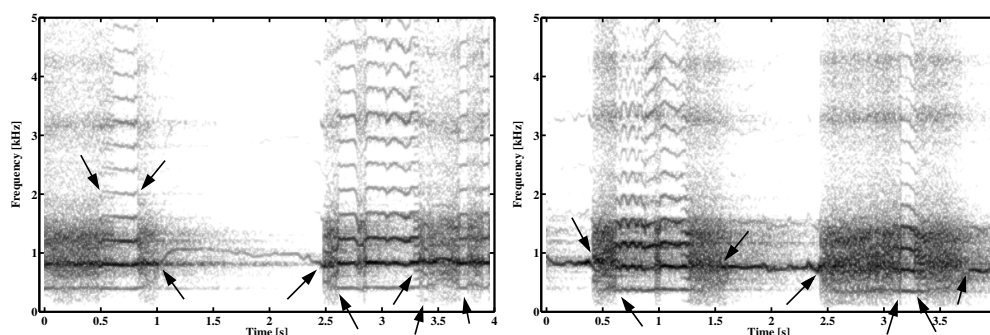


Figure 3.6: Spectral bifurcation diagrams showing recurrent instances of abrupt transitions from regular phonation to subharmonics and irregular phonation (arrows) within the same vocalization sequence. The samples were taken from a recording of the solo vocal improvisation *entre nosotros - epitafio a las ballenas* by the vocal improviser Fatima Miranda (Miranda, 1992).

There, stronger harmonic components than during the biphonic episode could be observed.

In Fig. 3.3 an example of a female voice is given revealing a sudden transition from normal, harmonic phonation to a voice sound with a broadband spectrum. Sequences of period doubling and quadrupling bifurcations appeared before the onset of irregular aperiodic oscillation. Within the irregular segment, starting at about 2.0s, residual spectral components could be seen. These were related to the previous harmonic components. Additionally, formants at approximately 1.0 kHz, 1.4 kHz and 3.0 kHz could be seen. The subharmonic components bifurcating at about 0.5s and approximately 1.3s were perceived as time varying mixtures of harmonic and inharmonic components.

In Figures 3.4 and 3.5 examples are presented where varying formant frequencies that matched harmonics coincided with bifurcations to various subharmonic regimes. Fig. 3.4 is the visualization of an extremely high and light fundamental frequency of a female voice at about 1.25 kHz modulated by a tremolo frequency of about 4 Hz. Sidebands of the second harmonic appeared when the upward moving formant frequency matched the second harmonic at about 1.0s. The result was to reinforce the amplitude of this harmonic. The sidebands of the fundamental frequency and the third harmonic were also visible, but lower in intensity than the sidebands of the second harmonic.

In Fig. 3.5 an example of a formant matching harmonics associated with a period multiplying bifurcation is shown. A male singer used the falsetto register to phonate at about 550 Hz. A formant frequency coincided with the fourth harmonic around 0s. At the same time period quintupling (P5) occurred. Then, from 0.5s to

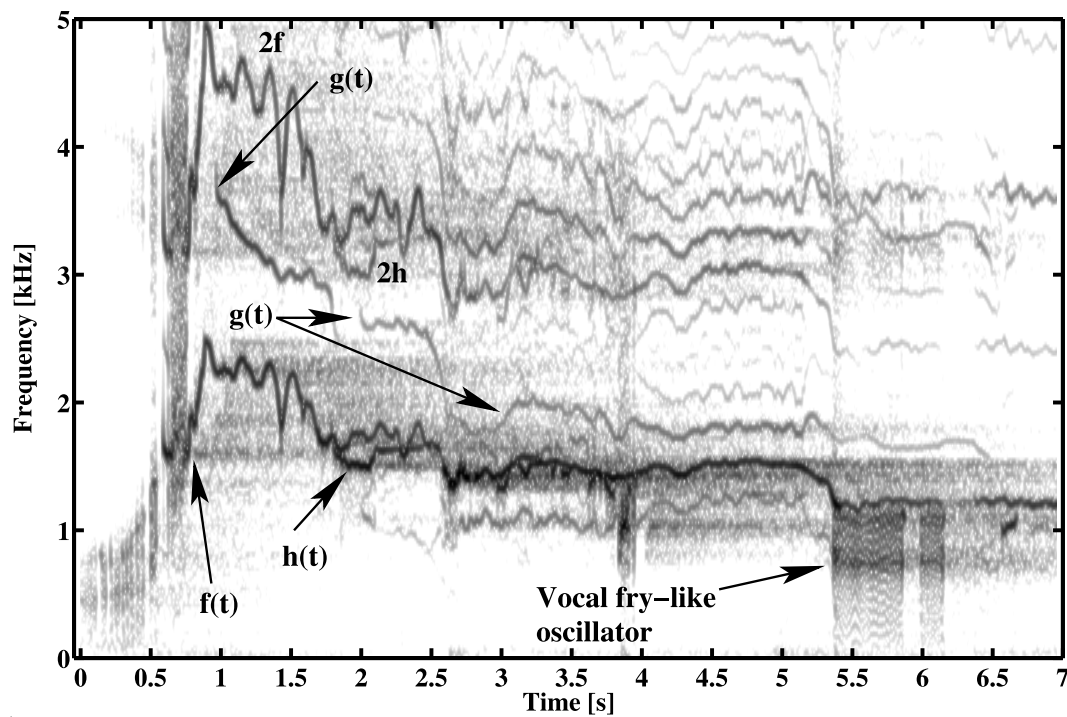


Figure 3.7: Spectral bifurcation diagram displaying a female glottal whistle with biphonation and triphonation. In the final segment triphonation is observed with two glottal whistle components and a third, vocal fry-like oscillator. The sample was taken from a recording of the solo vocal improvisation *Signals* by the vocal improviser Anna Homler (Homler, 1989).

3. Nonlinear Phenomena in Contemporary Vocal Music

1.5 s the formant frequency moved downward and coincided roughly with the third harmonic at about 1.5 s (arrow). At this point period quintupling (P5) occurred.

Fig. 3.6 shows recurrent instances of abrupt transitions from regular phonation to subharmonic and irregular phonation. The female voice produced phonation at nearly 900 Hz. The left part of Fig. 3.6 began with an irregular segment with a periodic window followed by another irregular segment. At about 1.0 s an extremely high and light phonation with an increased perception of pitch set in. Note that about 1.0 – 2.5 s the frequency component of 900 Hz was an artefact of the recording environment. It was the result of excess reverberation during recording, not representative of phonation. At about 2.5 s a transition to a chaotic segment with residual harmonics occurred followed by harmonic windows with subharmonics including P6 at 3.3 s. At about 3.4 s a chaotic segment with residual harmonics started, which was interrupted by short instances of subharmonic windows. In the right part of Fig. 3.6 the qualitative behavior was similar, although details varied. At about 0.4 s the high and light phonation bifurcated to a chaotic behavior. This segment was interrupted by a subharmonic window. After the chaotic segment the voice transitioned again to the high, light phonation. At about 2.5 s another chaotic segment started which, again, was interrupted by a subharmonic window at about 3.2 s.

In Fig. 3.7 a female “glottal whistle” is shown. It started at a pitch of above 2.0 kHz descending to about 1.5 kHz. After the initial “glottal whistle”, a second frequency of approx. 3.5 kHz appeared at about 1.0 s and was functionally independent. At 2.0 s a third frequency showed up which vanished at about 2.5 s. As the two remaining frequency contours descended further, multiple combination frequencies could be observed which increased in amplitude. This increase of intensity could be related to an increase of coupling (i.e. nonlinearities) over time between the two independent frequencies. At about 5.4 s a third component appeared which perceptually had a vocal fry-like character.

3.5 Discussion

This chapter shows examples of the intentional use of nonlinear phenomena in contemporary vocal music. In multi-tone complex sonorities of vocal improvisors subharmonics, biphonation, sudden onset of irregular phonation (possibly deterministic chaos) interspersed with more periodic windows and sudden frequency jumps (register changes) can be found. This study provides arguments that nonlinear phenomena, well-known in voice pathology, play an essential role in modern artistic vocalizations. In the following sections bifurcation analyses, potential physiological mechanisms, issues relative to reproducibility, and their musical relevance will be discussed.

3.5.1 Bifurcation analysis

According to the theory of nonlinear dynamics all systems feature a limited set of qualitatively different dynamical behaviors. This study documents how this applies to sound production of the voice, in particular for solo vocal improvisations by performers using their voices for artistic sound production. The attractors relevant for the physical description of oscillatory systems are limit cycle (related to periodic oscillation), folded limit cycle (related to subharmonics), torus (related to two frequency oscillation) and chaos (related to highly irregular, aperiodic and noise-like behavior).

Transitions between different dynamic regimes (i.e. different attractors) are predicted to occur even for slowly varying system parameters such as pitch, formants, subglottal pressure or the varying interaction with supraglottal tissue structures. A comprehensive visualization of transitions can be achieved by bifurcation diagrams (Glass and Mackey, 1988) which display different dynamical behavior depending on one or two varying system parameters. Such diagrams were calculated for a simplified two-mass model of vocal folds (Steinecke and Herzel, 1995) and continuum models (Berry et al., 1994). Furthermore, bifurcation diagrams were measured for excised larynx experiments (Berry et al., 1996), and analyzed for a voice with unilateral paralysis (Mergell et al., 2000).

In this study time-frequency plots are used (spectrograms) as spectral bifurcation diagrams (Lauterborn and Suchla, 1984; Lauterborn, 1986; Lauterborn and Parlitz, 1988). In contrast to bifurcation diagrams obtained from mathematical models or experimental setups where single system parameters can be varied, there are no direct measurements of the varying system parameters in *in vivo* multi-tone complex sonorities of vocal improvisors. Thus in spectral bifurcation diagrams time is considered as a “control parameter” of the system.

There are a few hints on changing parameters such as pitch, formants, and subglottal pressure due to different shading of spectral lines and noise-like elements in the spectrographic displays. Thus it can be speculated about the physiological mechanisms that determine the vocal outputs.

3.5.2 Physiological mechanisms

This study suggests that formant induced transitions might play a role in the production of subharmonic and biphonic sequences. In Fig. 3.4 an example is shown for a formant induced transition of a harmonic phonation to an oscillation with sidebands. Thus, a second frequency modulates the initial harmonic output during formant matching which can be observed as sidebands in the spectrogram. With respect to nonlinear dynamics nomenclature, such transitions might be called secondary Hopf bifurcations which change the behavior of the system from a limit

3. Nonlinear Phenomena in Contemporary Vocal Music

cycle oscillation to a toroidal oscillation, a torus in phase space. The dominant formant, varying over time, seems to play a crucial role in destabilizing the glottal oscillator coupled to the vocal tract. Similarly, in Fig. 3.5 an example is given for a formant induced bifurcation of a harmonic vocal output to subharmonic oscillation. This conclusion is supported by previous model studies in simplified models for source-tract interaction. Sub- and supraglottal resonances were found to be relevant for high-pitched phonation and, moreover, biphonation (Mergell and Herzel, 1997b).

Second, it has been reported that sub- and supraglottal oscillators (and resonance factors) such as ventricular fold phonation, epiglottic and arytenoid cartilage can contribute to a variety of complex modes of phonation (Švec et al., 2000).

Third, another mechanism that can induce transitions of dynamical behavior is vocal fold asymmetry. Left-right asymmetries and anterior-posterior asymmetries facilitate subharmonic oscillation, biphonation and even chaotic vibration (Tigges et al., 1997; Ward et al., 1969; Neubauer et al., 2001; Berry et al., 1996).

Fourth, excessive subglottal pressure, increased stiffness of the vocal fold mucosa, or reduced prephonatory glottal shape (phonation neutral area) leading to chaotic vibrations have been identified in a recent study of a simplified two-mass model, even for a symmetrical vocal fold configuration (Jiang et al., 2001b). Excessively high airflow with a lax laryngeal posture has been implicated in the simultaneous production of subharmonics and deterministic chaos (Edgerton et al., 2001). It is argued that vocal improvisors exploring the full range of vocal abilities are able to use exceptional, but not pathological methods, to induce multi-tone complex vocal sonorities such as (multiple) subharmonics, biphonation and irregular aperiodic sound signals. (e.g., see Fig. 3.3 and 3.6).

Fifth, it is argued, that vocal improvisors can use glottal whistles to produce complex sonorities. The term “glottal whistle” describes vortex-generated sound with partially adducted glottal or supraglottal constrictions that additionally may feature slight vocal fold vibrations (vortex-induced fold vibrations). This study suggests that these biomechanic or aerodynamic vibrations may be produced at different locations simultaneously within the glottis. The subglottal airstream as the driving force could facilitate one possible coupling between these different oscillators. Thus, biphonation or triphonation could occur due to several interacting oscillators: left versus right vocal fold, anterior versus posterior parts of the vocal folds, glottal versus supraglottal vibrating structures, or vortex-induced vibrations. Moreover, as typical formant frequencies lie in the range of several kHz, they could thus match the pitch of the glottal whistle, and either could stabilize biphonation or induce transitions to oscillations with sidebands or even irregular behavior (Edgerton et al., 2001). The highly variable time course (melody) shown in Fig. 3.7 suggests the lack of accurate control mechanisms. The melody may

3. Nonlinear Phenomena in Contemporary Vocal Music

thus result from pressure/flow regulated behavior or from relaxation behavior of tissue stiffness.

3.5.3 Reproducibility, intention, and control

Resulting from personal performance experiences, the development of special training programs attended by amateur and professional voices, and the acoustic analysis of hundreds of sound samples of vocal improvisors it can be claimed that singers are able to use nonlinear phenomena intentionally. This study argues that the above mentioned physiological mechanisms ensure reproducibility in a musical context. These mechanisms include vocal tract resonance tuning, subglottal pressure adjustments, or voluntary desynchronization of vibration modes using asymmetries. Although minute details of bifurcations to subharmonics, biphonic or irregular segments vary (e.g., see Fig. 3.5 and 3.6) the vocal sonorities can be produced in artistically controlled methods. In contrast to pathological phonation where nonlinear phenomena occur accidentally, certain vocalists have the control of recurrent production of complex sonorities. Musically, the control aspect is important for the sound output. By changing perceptually robust variables over time an intended affect, emotion or meaning can be produced (Sachs, 1962; Mazo, 1994). Next, as many performance techniques of vocal improvisors may last the duration of a performer's career or even life-span, it is obvious that these techniques are not necessarily aberrant behaviors to be avoided (Levin, 1996), but rather can be physically and spiritually beneficial (Rachele, 1997; Titze, 1999).

3.5.4 Musical relevance

The insight provided by nonlinear dynamics to the understanding of complex musical vocalizations can offer valuable information about an instrument with no levers, buttons or strings – the voice. This type of analysis offers the identification of a phenomenon with a class and a process. For example, biphonation is identified as a torus that appears via a secondary Hopf bifurcation.

One readily-available tool relevant for sound analysis is the spectrogram. In the context of acoustic chaos spectrograms are called spectral bifurcation diagrams (Lauterborn and Suchla, 1984). When time is considered as a control parameter, they are also termed “visible speech” (Lauterborn and Parlitz, 1988). Spectrograms may also contribute to the training and treatment of desired or aberrant voice signals as windows of visualization. When combined with information regarding airflow, laryngeal posturing and tension, etc. this study proposes to use them similar to instrumental fingering charts (standard pedagogical tools for producing nonstandard extended complex instrumental sonorities (e.g., Kimura, 1995)).

Chapter 4

Spatio-temporal Analysis of Irregular Vocal Fold Oscillations: Biphonation due to Desynchronization of Spatial Modes

This chapter is a revised version of Neubauer, J., Mergell, P., Eysholdt, U., and Herzel, H. (2001), “Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes,” J. Acoust. Soc. Am. 110, 3179–3192. (Neubauer et al., 2001)

This chapter is a report on direct observation and modal analysis of irregular spatio-temporal vibration patterns of vocal fold pathologies *in vivo*. The observed oscillation patterns are described quantitatively with multi-line kymograms, spectral analysis, and spatio-temporal plots. The complex spatio-temporal vibration pattern are decomposed by empirical orthogonal functions into independent vibratory modes. It is shown quantitatively that biphonation can be induced either by left-right asymmetry or by desynchronized anterior-posterior vibratory modes, and introduce the term “AP (anterior-posterior) biphonation”. The presented phonation examples show that for normal phonation the first two modes sufficiently explain the glottal dynamics. The spatio-temporal oscillation pattern associated with biphonation due to left-right asymmetry can be explained by the first three modes. Higher order modes are required to describe the pattern for biphonation induced by anterior-

posterior vibrations. Spatial irregularity is quantified by an entropy measure, that is significantly higher for irregular phonation than for normal phonation. Two asymmetry measures are introduced, the left-right asymmetry and the anterior-posterior asymmetry, as the ratios of the fundamental frequencies of left and right vocal fold and of anterior-posterior modes, respectively. These quantities clearly differentiate between left-right biphonation and anterior-posterior biphonation. This study proposes methods to analyze quantitatively irregular vocal fold contour patterns *in vivo* and complements previous findings of desynchronization of vibration modes in computer models and in *in vitro* experiments.

4.1 Introduction

In the last years, several new promising concepts in the fields of voice and phonation diagnostics, modeling, signal classification and analysis have been developed to shed light on mechanisms that cause irregular vocal fold oscillations. Laryngeal (video) stroboscopy is widely used in everyday clinical examinations as a tool for visualizing pathological vocal fold dynamics, but it is only appropriate for the investigation of periodically vibrating vocal folds (Wendler et al., 1996). On the contrary digital high-speed glottography allows the direct observation of the glottal dynamics and the separate analysis of the left and the right vocal fold vibrations (Eysholdt et al., 1996; Kiritani et al., 1993; Hammarberg, 1995). This examination method allows to record data required for the analysis of transient, subharmonic and temporally and spatially aperiodic vocal fold dynamics. Digital image processing algorithms can be applied to obtain time series of the oscillations of the vocal fold edges for further analysis (Wittenberg, 1998).

The theory of nonlinear dynamics provides the framework for classifying vocal instabilities (Herzel, 1993; Titze et al., 1993). Stationary oscillations can be related to low dimensional attractors (limit cycle, torus, chaotic attractor) (Bergé et al., 1984). Qualitative changes of the vocal fold dynamics due to variations of the myoelastic and aerodynamic properties can be classified as bifurcations. Examples are the phonation onset (Hopf bifurcation) and the appearance of subharmonics (period doubling bifurcation). In several studies, attractors and bifurcations have been analyzed in biomechanical models of the vocal folds (Berry et al., 1994; Lucero and Gotoh, 1993; Steinecke and Herzel, 1995; Mergell, 1998). Recently, vocal irregularities have been described by combining digital high-speed glottography and biomechanical modeling (Mergell et al., 2000). The observed time series from a pathological voice could be reproduced with quantitative agreement in a simplified two-mass model by parameter adjustment. The underlying model as-

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

sumptions neglect changes of the vocal fold properties along the anterior-posterior (AP) direction. Therefore, the model only captures left-right (LR) asymmetries.

However, there are several indications that anterior-posterior (AP) oscillation modes can contribute to irregular glottal oscillation patterns (Farnsworth, 1940; Titze, 1973; Berry et al., 1994; Hess et al., 1994; Tigges et al., 1999; Švec et al., 2000; Berry, 2001). Berry et al. (1994) used empirical orthogonal functions (EOFs) to determine the dominant spatial modes in a finite element model of the vocal folds. They showed that even complicated vibration patterns can be explained by a few modes. Moreover, higher-order modes could be extracted from this biomechanical model for one case of irregular phonation. EOF studies of vocal fold tissue (e.g., excised larynx experiments) confirmed and extended the results obtained from theoretical models (Berry, 2001). Resonance studies of human vocal folds *in vivo* showed that the resonance frequencies correspond to distinct anterior-posterior modes (Švec et al., 2000). Resonance frequencies of the modes of a finite element model were found to be grouped around resonance frequencies that correspond to different anterior-posterior modes (Berry, 2001). From an endoscopic view, the modes within these groups cannot be distinguished, as they exhibit the same anterior-posterior vibration pattern. Generally, inferior-superior vibration modes cannot be resolved from the endoscopic view because the upper vocal fold edge masks the lower parts of the folds.

Models more complex than simplified two-mass models are required to reproduce irregular spatio-temporal vibration patterns (Titze, 1973; Titze and Strong, 1975; Titze, 1976; Alipour-Haghighi et al., 2000). Since there is always a tradeoff between the spatial accuracy of measured myoelastic properties and the physiological completeness of the model, it is important to know how complex a low-order model must be to describe the vocal fold dynamics *in vivo* accurately. Modal analysis can answer this question and is an appropriate tool for the design of low-order models. The number of dominant empirical modes is related to the number of horizontal and vertical degrees of freedom that biomechanical models must provide to capture the effective glottal dynamics (Aubry et al., 1991; Breuer and Sirovich, 1991; Berry et al., 1994). The concept of modal analysis has not yet been applied for spatio-temporal analysis of glottal vibration patterns *in vivo*.

This chapter focuses on the quantitative study of anterior-posterior modes and of irregular phonation by desynchronization of anterior-posterior modes of vibration. Thus, in the following sections a quantitative analysis of complex and irregular spatio-temporal vibration patterns of the vocal folds is presented based on three phonation examples. One example with normal healthy phonation is analyzed that serves as a reference. In addition, two high speed sequences with irregular phonation are chosen to demonstrate two different kinds of mechanisms that lead to biphonation. The term “biphonation” characterizes phonation with

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

two fundamental frequencies (Mergell and Herzel, 1997b; Wilden et al., 1998; Kiritani et al., 1995; Ishizaka and Isshiki, 1976; Herzel and Wendler, 1991; Herzel and Reuter, 1997; Smith et al., 1992). In general, biphonation can be induced by different glottal, pharyngeal, and aerodynamical oscillators with different oscillation frequencies. Possible mechanisms for biphonation are the natural or pathological asymmetry of vocal folds, the decoupling of vertical and horizontal vibratory modes within single vocal folds, the interaction with other mechanical or aerodynamical oscillators in the sub- and supraglottal airway, and the interaction with vortices produced by the glottal constriction of the airstream. (Herzel and Reuter, 1997; Mergell and Herzel, 1997b; Tigges et al., 1997)

As a first image processing step the technique of multi-line kymography is introduced to extract time series of glottal edge points from digital high-speed recordings. Spectral analysis and spatio-temporal plots of the multi-line time series are used as conventional analysis tools both for simple data analysis and consistency check. Furthermore, empirical orthogonal eigenfunctions (EOFs) are computed from the time series of the vocal fold edge displacements about the estimated glottal midline. It is argued that EOF analysis is the appropriate way to deal with spatio-temporal vibration patterns. Thus, the multi-line time series are decomposed into empirical orthogonal functions. This method is used to measure the degree of desynchronization of anterior-posterior modes.

Furthermore, four different measures are suggested to quantify objectively spatio-temporal vibration patterns. The asymmetry coefficient according to Mergell et al. (2000) is used to measure left-right asymmetry. This coefficient is related to the asymmetry coefficient in simple two-mass models (Steinecke and Herzel, 1995). Similarly, an anterior-posterior asymmetry measure is introduced to describe anterior-posterior mode desynchronization. An entropy measure is used to quantify spatial irregularity, and the number of dynamically relevant glottal modes are estimated.

4.2 Material and Methods

Digital high-speed image sequences during phonation were recorded with the camera system CAMSYS+ 128 (described by Bloss et al., 1993) at a sampling rate of 3704 frames per second with a spatial resolution of 128×64 pixels in combination with a rigid larynx endoscope (for further technical details and caveats cf. Wittenberg et al. (1995) and Wittenberg (1998)). The high-speed data of three subjects were examined: subject JN (male, 27 years) with a healthy voice, subject WS (female, 29 years) with a left recurrent nerve paralysis, and subject MM (female, 26 years) with a functional dysphonia. On basis of the Roughness-Breathiness-Hoarseness (RBH) scale subject WS was judged to be R1B1H2. Clinical stro-

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

boscopy revealed a prephonatory standstill of the left vocal fold, whereas the right vocal fold was found to be normal. During phonation stroboscopy was not applicable due to biphonic oscillation of the vocal folds. Clinical standard examination of subject MM gave no visible morphological peculiarities. The voice was judged on basis of the Roughness-Breathiness-Hoarseness (RBH) scale to be R2B1H2 with diplophonic episodes. In the following study, the high-speed sequences were labeled “normal voice”, “recurrent nerve paralysis” and “functional dysphonia”. These particular pathological phonations were chosen, because preliminary analysis of the corresponding sound data indicated biphonation. The pathological phonation examples were chosen out of about 100 recordings with indications for biphonation out of overall 4000 high-speed sequences. About 10 of the biphonic phonations showed anterior-posterior vibration patterns. Figures 4.1–4.3 show one oscillation cycle taken from the digital image sequences of each subject. The highlighted glottal edge contours (i.e. the entirety of all glottal edge points of each vocal fold) were the result of the image processing algorithm explained below. The algorithm for data reduction was a simplified and modified version of the kymographic image processing proposed by Wittenberg (1998).

4.2.1 Multi-line kymography

As a first processing step, all horizontal scan lines of each digital image were extracted over the visible vocal fold length, i.e. from the posterior side (arytenoid cartilage) to the anterior side (thyroid cartilage). The arytenoid side corresponds to the upper part of the digital images shown in Figs 4.1–4.3, whereas the thyroid side corresponds to the lower part. The extracted scan lines were subsequently concatenated. In Figs 4.4–4.6 the vocal fold dynamics is visualized by a few hundred subsequent scan lines. The resulting gray scale arrays (kymograms) showed the change in distance between edge points on left and right vocal folds during phonation (Švec and Schutte, 1996; Tigges et al., 1999). The tracked glottal edge points of the left (upper line) and the right vocal fold (lower line) were marked by white points.

These edge points were determined with a fixed gray value threshold for each kymogram. This threshold separated between image points from the glottal aperture and from the vocal fold tissue. As a rigid endoscope was used viewing the vocal folds from above (superior view), the tracked points corresponded to points of the superior medial vocal fold edge or the inferior vocal fold edge. During the opening phase the tracked points were close to the superior medial edge, whereas during the closing phase the tracked points corresponded to a point between superior and inferior edge. However, due to the limited spatial resolution of the camera system vertical vibration modes of the vocal fold cover could not be resolved. The movement of the vocal fold edge points was related to the oscillations riding on the

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

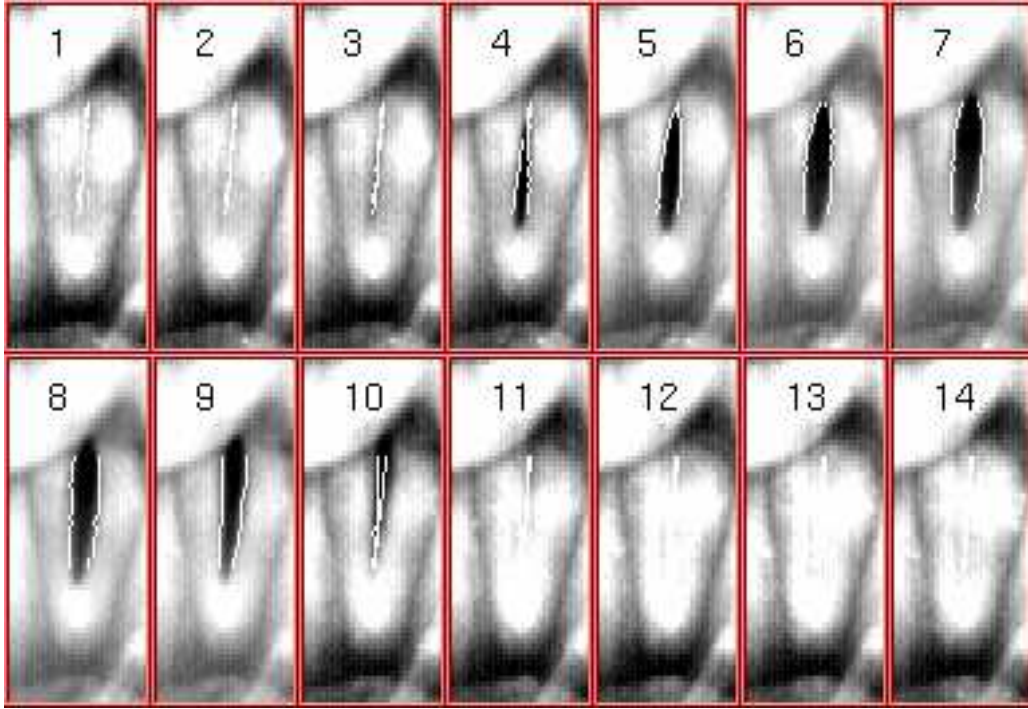


Figure 4.1: Digital high-speed image sequence showing a glottal cycle with 21 tracked glottal contour points per vocal fold: subject JN with normal phonation. Every second frame is shown, the time interval between successively plotted frames is $\Delta t = 0.54$ ms. Note that the right arytenoid cartilage is visible in the left upper corner of the images. Thus, the upper and lower side of each frame correspond to the posterior and anterior side of the vocal folds, respectively. The left and right vocal folds are displayed on the right and left side of the digital images, respectively.

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

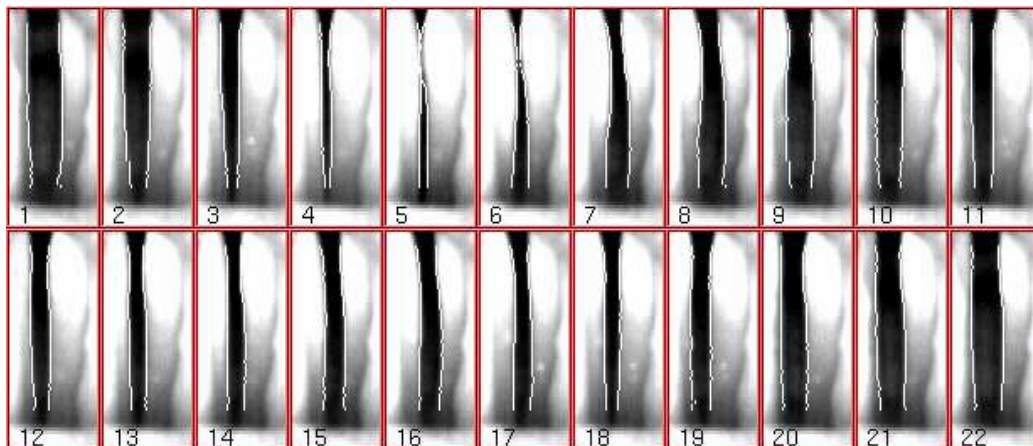


Figure 4.2: Digital high-speed image sequence with 49 tracked glottal contour points per vocal fold of subject WS paralysis (every second frame shown, $\Delta t = 0.54$ ms for plotted frames). The upper and lower side of each frame correspond to the posterior and anterior side of the vocal folds, respectively.

black-white interface of the binary segmented multi-line kymograms (Wittenberg, 1997; Mergell et al., 2000). The extracted time series were termed “high-speed glottograms” (HGGs) following Wittenberg et al. (1995). Table 4.1 gives details on the data sets of the different phonation examples used for spatio-temporal analysis of vocal fold vibration.

4.2.2 Time series preprocessing

The HGG data had to be preprocessed to avoid artifacts caused by relative horizontal movements and rotations of the observing endoscope about the glottis. Following previous work (Mergell et al., 2000), just horizontal movements were corrected neglecting vertical movements that were assumed to be small within the short selected HGG sequences (Wittenberg, 1998). Therefore, the displacements of vocal fold edges about the glottal midline were calculated, which was estimated frame by frame (Fig. 4.7).

First, the glottal midline was approximated by a regression line through the midpoints $(\bar{x}_k(t_i), y_k(t_i))$ of each scan line k , where the midpoints were:

$$\bar{x}_k(t_i) = \frac{1}{2} \left(x_k^{(left)}(t_i) + x_k^{(right)}(t_i) \right). \quad (4.1)$$

Note that $y_k(t_i) = y_k^{(left)}(t_i) = y_k^{(right)}(t_i)$. Here, $k = 1, \dots, N$ indicated the

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

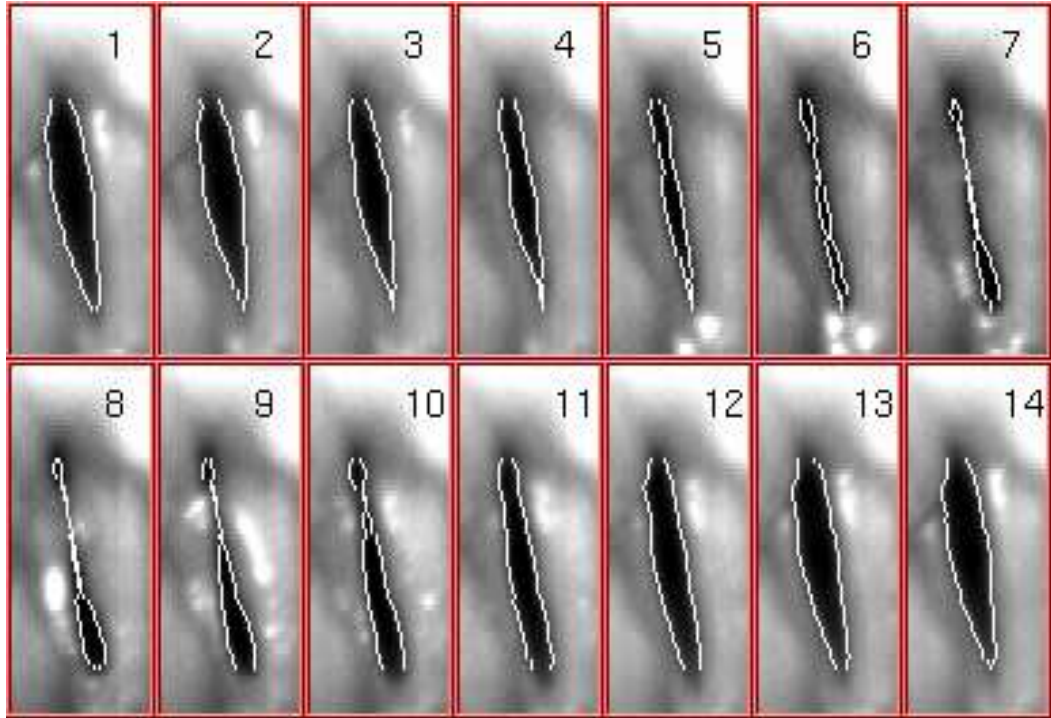


Figure 4.3: Digital high-speed sequence with 39 tracked glottal contour points per vocal fold of subject MM (every frame shown, $\Delta t = 0.27$ ms). The upper and lower side of each frame correspond to the posterior and anterior side of the vocal folds, respectively.

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

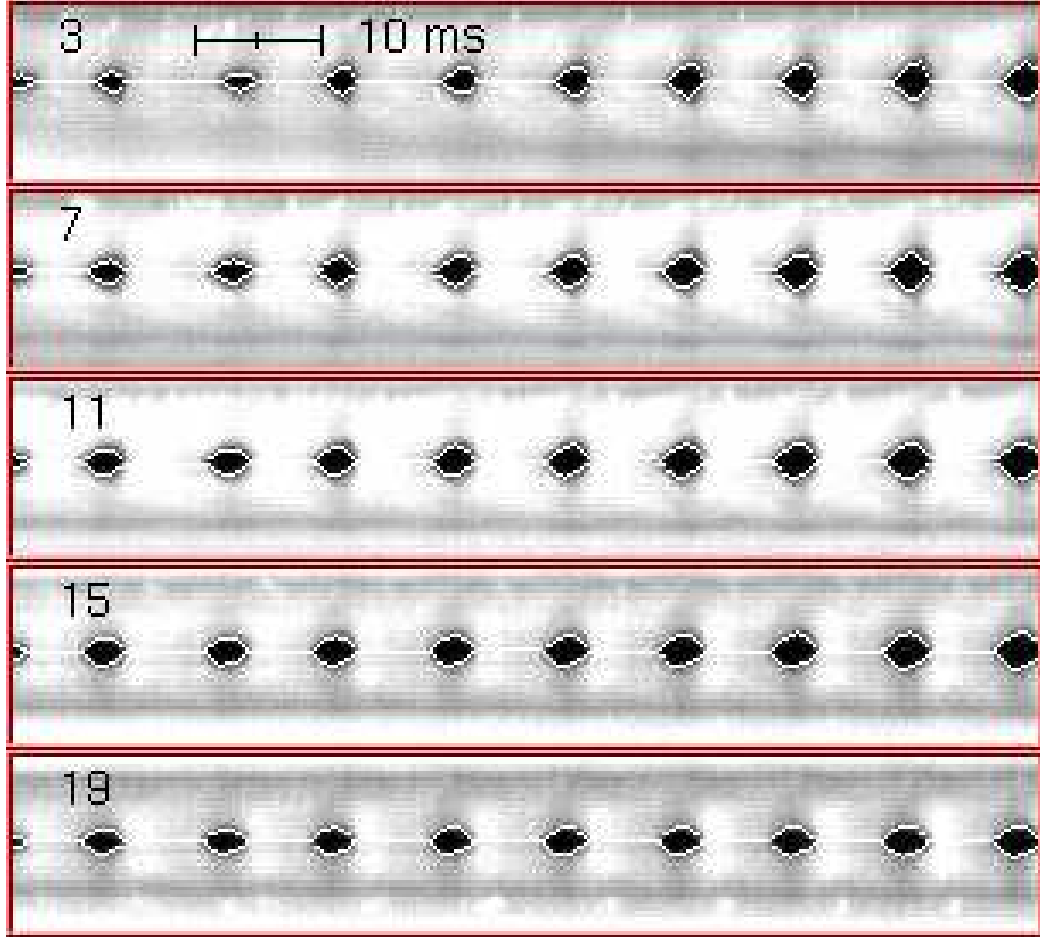


Figure 4.4: Selected multi-line kymograms (chosen from 21 kymograms) with highlighted extracted time series for subject JN (with normal phonation). The upper white line corresponds to the left vocal fold edge, the lower white line to the right vocal fold edge. Length of shown kymograms is $T = 81$ ms. The numbers correspond to the scan line number along the posterior-anterior direction. The glottal aperture is black-coded, the surrounding is shown in different gray values. The vertical direction within each kymogram corresponds to the left-right direction of the digital high-speed image frames.

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

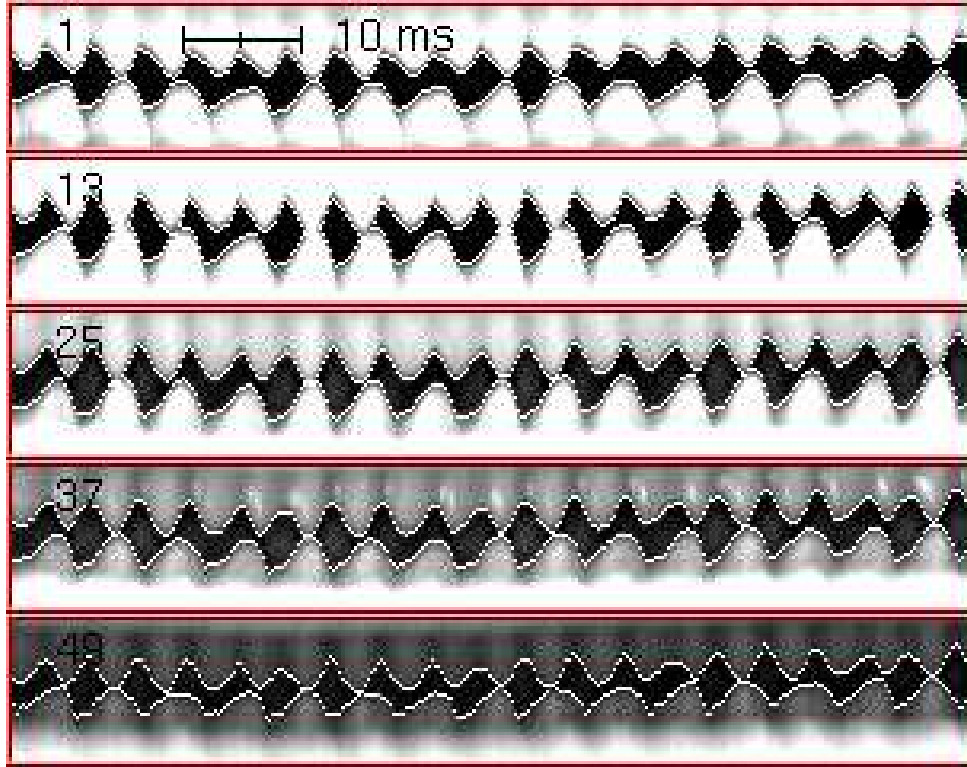


Figure 4.5: Selected multi-line kymograms for subject WS. Length of shown kymograms is $T = 81$ ms. The upper white line follows the left vocal fold edge, the lower white line follows the right vocal fold edge. In each kymogram desynchronization of the left and right vocal fold oscillation is observed: Five oscillation maxima of the left vocal fold (upper line) correspond to four oscillation maxima of the right vocal fold (lower line).

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

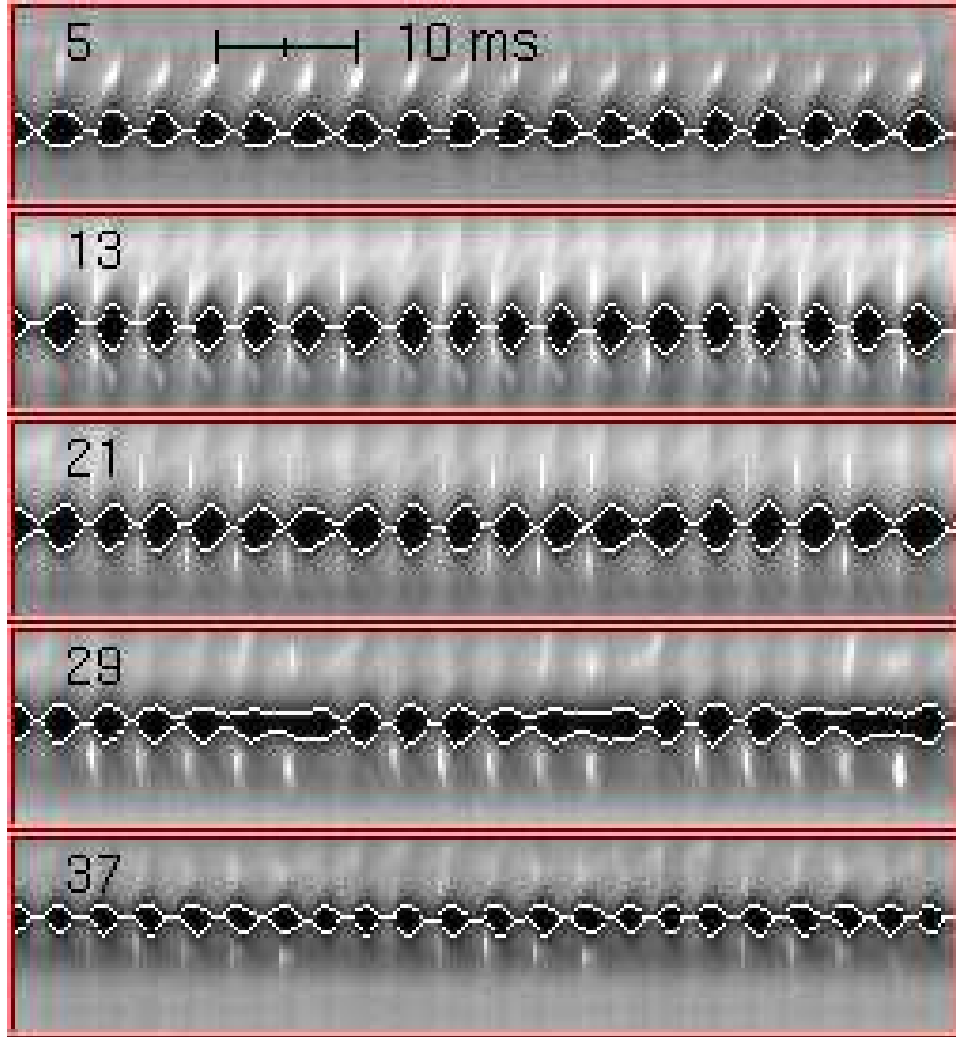


Figure 4.6: Selected multi-line kymograms for subject MM. Length of shown kymograms is $T = 67$ ms. The upper white line indicates the left vocal fold edge, the lower white line indicates the right vocal fold edge. In each kymogram nearly symmetrical oscillations of both vocal folds are observed. However, an increasing modulation of both vocal fold oscillations can be seen comparing kymograms 13, 21 and 29 along the posterior-anterior direction. Thus, desynchronization of vocal fold oscillations along the anterior-posterior direction is observed.

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

Subject	Phonation	Number of glottal contour points	Length of time series [ms]	Typical frequency [Hz]
JN (healthy)	normal	21	1215	120
WS (recurrent nerve paralysis)	irregular	49	618	197
MM (functional dysphonia)	irregular	39	354	270

Table 4.1: Details on time series from different subjects with normal and irregular phonation. The sampling rate was 3704 frames per second for all time series. The difference in length is due to the different length of high-speed sequences obtained from the clinical investigation and due to the selection of stationary segments. With respect to the typical oscillation frequencies the number of measured cycles are sufficient for a statistical analysis such as the EOF estimation calculation.

kymogram number and $i = 1, \dots, M$ was the time index (see Fig. 4.7). The estimated glottal midline of frame t_i was described by the slope $m(t_i)$ and the intercept $b(t_i)$. The slope and the intercept both contained parts of the glottal dynamics and slow modulations caused by relative movements between the high-speed camera and the glottis. As the HGG data were only corrected for slow modulations, a moving average procedure was applied to the time series $m(t_i)$ and $b(t_i)$ (length of averaging window: 50 ms, corresponding cutoff frequency: 20 Hz). The cutoff frequency of this low-pass filtering process was chosen close to the maximum tremor frequency of about 16 Hz (Riviere et al., 1998). Thus, the filtered values $\bar{m}(t_i)$ and $\bar{b}(t_i)$ were slowly modulated. The vocal fold displacements $d_k^{(\alpha)}(t_i)$ were calculated as the distance of the extracted contour points from the estimated glottal midline:

$$d_k^{(\alpha)}(t_i) = \left[x_k^{(\alpha)}(t_i) - \frac{y_k(t_i) - \bar{b}(t_i)}{\bar{m}(t_i)} \right] \sin(\varphi) . \quad (4.2)$$

Here, $\varphi = \arctan(\bar{m}(t_i))$ was the inclination angle of the glottal midline, $\alpha \in \{left, right\}$ specified the side, and $x_k^{(\alpha)}(t_i)$ and $y_k(t_i)$ were the image coordinates

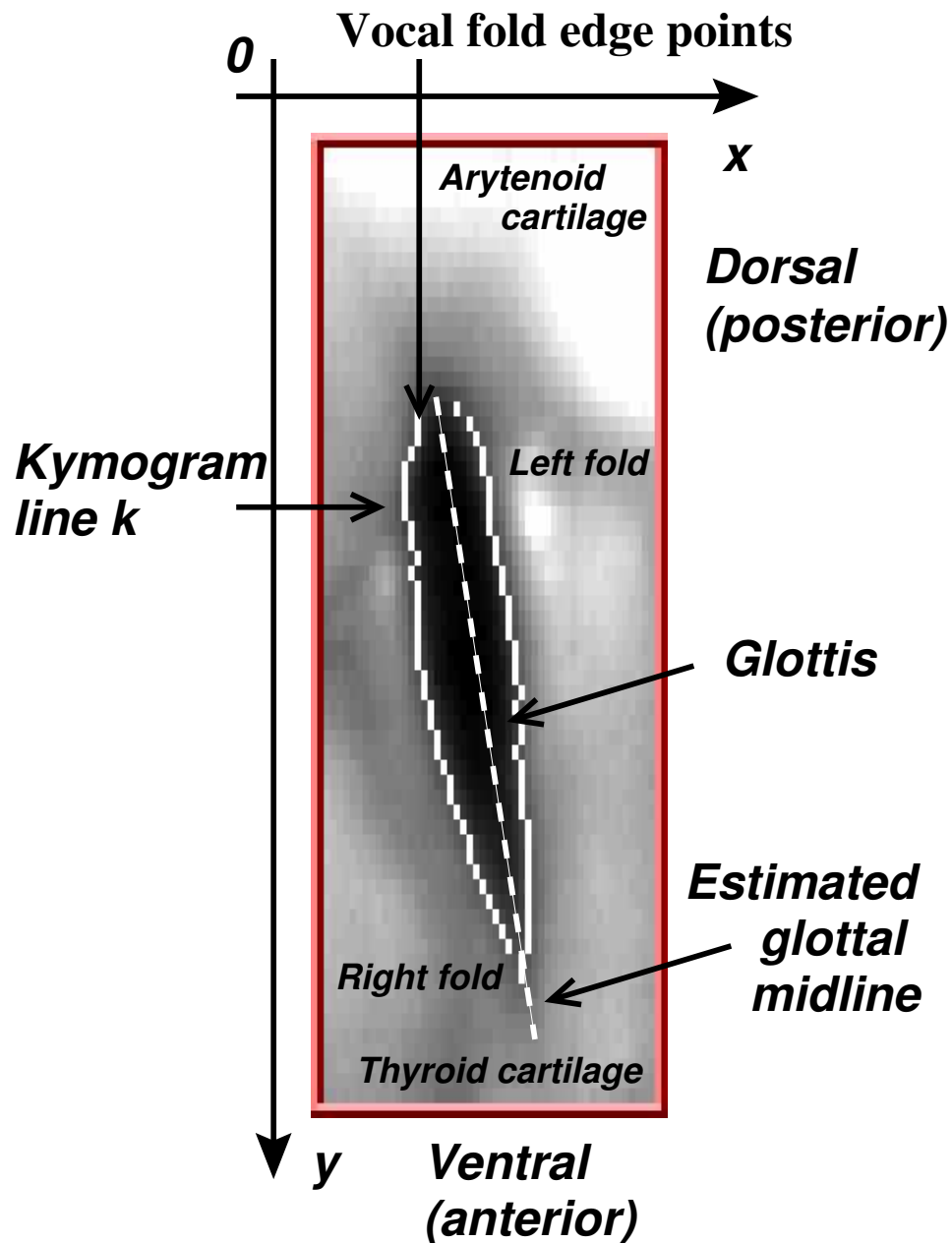


Figure 4.7: Coordinate system for analysis of digital high-speed images: Vocal fold edge points of left and right vocal fold are highlighted. The distances of the estimated glottal midline to the vocal fold edge points are used as vocal fold displacements for further analysis.

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

of the vocal fold contour points (see Fig. 4.7). In this way, slow horizontal and rotational modulations contained in $\bar{m}(t_i)$ and $\bar{b}(t_i)$ were removed from the HGG data $(x_k(t_i), y_k(t_i))$.

4.2.3 Calculation of the empirical orthogonal functions

The technique of empirical orthogonal functions is an appropriate method for analyzing complex and irregular spatio-temporal glottal motion patterns. It can be used to decompose glottal contour dynamics into principal modes of vibration (Berry et al., 1994).

As only the oscillation of the vocal folds about their dynamic rest position was of interest, the temporal averages (average over whole length of time series) were subtracted from the vocal fold displacements of left and right vocal folds:

$$\delta_k^{(\alpha)}(t_i) = d_k^{(\alpha)}(t_i) - \bar{d}_k^{(\alpha)}. \quad (4.3)$$

The EOFs $\Phi_k^{(\alpha)}$, $k = 1, \dots, N$, were calculated from the oscillatory components $\delta_k^{(\alpha)}(t_i)$. EOFs represent a decomposition of the contour dynamics $\delta_k^{(\alpha)}(t_i)$ into different spatial oscillating modes.

Furthermore, the corresponding eigenvalues $\lambda_k^{2(\alpha)}$ were computed. They were the weights or the contributions of the EOFs on average to the complete spatial dynamics.

To analyze the dynamics of the EOFs, the time-varying contribution of different modes (EOFs) $\Phi_k^{(\alpha)}$ were calculated. These contributions were called temporal coefficients $\Psi_l^{(\alpha)}(t_i)$. For a more detailed description of the empirical orthogonal functions, see Appendix A, page 179.

According to Mergell et al. (2000), the ratio $Q_{lr}^{(exp)}$ of the lower and higher fundamental frequency of left and right vocal folds parameterizes the laryngeal left-right asymmetry. Similarly, anterior-posterior asymmetry was measured with the ratio $Q_{ap}^{(exp)}$ of the lower and higher fundamental frequency of the first two temporal coefficients. This quantity measured the average temporal irregularity of the spatio-temporal oscillation pattern of a single vocal fold over the chosen time interval.

To parameterize the spatial irregularity Shannon's entropy $S_{tot}^{(\alpha)}$ was calculated from the EOF weights of each vocal fold. This entropy quantified the degree of disorder of the mode decomposition. For a further discussion of the entropy measure, see Appendix A, page 179.

For the sake of comparability and robustness, a simple threshold criterion was introduced for the estimation of dynamically relevant modes. The observed time series of the glottal contour points were reconstructed with a subset of the calculated EOFs. To find the required subset size, one starts with the first EOF with

the largest weight, adds subsequent EOFs with smaller and smaller values, until the difference between the reconstructed and the observed time series $\delta_k^{(\alpha)}(t_i)$ was smaller than a certain threshold.

4.3 Analysis of spatio-temporal glottal patterns

4.3.1 Spectral analysis of spatio-temporal plots

Spatio-temporal plots of the oscillatory component $\delta_k^{(\alpha)}(t_i)$ were used to discuss qualitatively the vibration patterns of different phonation examples. For this purpose, the time series $\delta_k^{(\alpha)}(t_i)$ was encoded in gray scale values in a space-time coordinate system. Bright regions corresponded to positive values, whereas dark areas indicated negative values. For the sake of visualization, the values of $\delta_k^{(\alpha)}(t_i)$ were normalized.

In Fig. 4.8 the desynchronization of left and right vocal fold vibrations for subject WS is illustrated. During four oscillation maxima on the right vocal fold, five oscillation maxima on the left vocal fold were observed. Left-right asymmetry was $Q_{lr}^{(exp)} \approx 4/5$. Note that there was no significant difference between the lower (posterior side) and the upper (anterior side) part of each spatio-temporal plot. Therefore, anterior-posterior asymmetry was estimated as $Q_{ap}^{(exp)} = 1$. Here, anterior-posterior asymmetry was determined as the ratio of lower and higher oscillation frequencies of vibrations close to the anterior or posterior end of the vocal folds.

The spatio-temporal plots for time series of subject MM (Fig. 4.9) showed nearly seven oscillation maxima in the upper part (anterior side) in contrast to nearly six maxima in the lower part (posterior side) of the vocal folds. Left-right symmetrical oscillations of the vocal folds, $Q_{lr}^{(exp)} = 1$ were present.

On the level of time series from single kymograms, spectra of selected time series of vocal fold edge points (HGGs) were examined. The spectra from subject WS showed a pronounced left-right asymmetry. In Fig. 4.10 two spectra for two different locations on each vocal fold are shown as examples. One location was close to the anterior side (kymogram line number $k = 40$), the other was close to the posterior side of the vocal folds (kymogram line number $k = 10$). All apparent peak frequencies $f_{m,n}$ could be explained by linear combinations of just two independent frequencies f_l and f_r : $f_{(m,n)} = mf_r + nf_l$, $m, n \in \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$. The fundamental frequencies were $f_r \approx 197$ Hz and $f_l \approx 265$ Hz. Thus, left-right asymmetry was $Q_{lr}^{(exp)} = f_r/f_l \approx 0.74$. For both anterior and posterior positions on the vocal folds, the corresponding spectra of left and right vocal folds revealed a significant left-right asymmetry. The peak frequency at f_r was dominant in the spectra of the right vocal fold, whereas for the left vocal fold the dominant

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

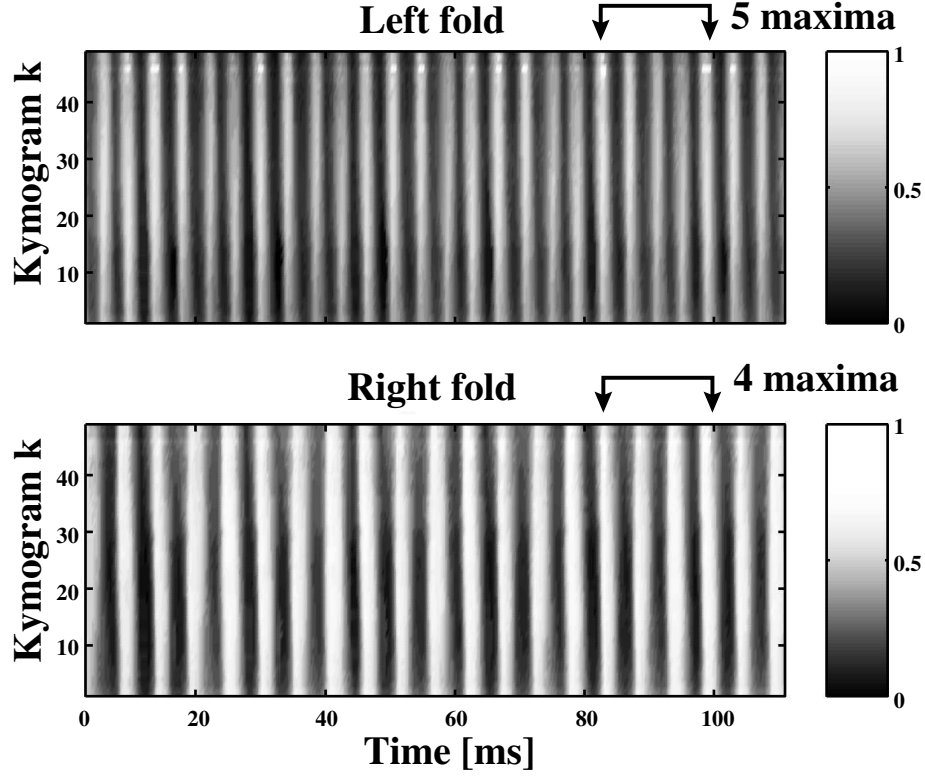


Figure 4.8: Spatio-temporal plots of the scaled oscillatory component $\delta_k^{(\alpha)}(t_i)$ of vocal fold edge points for subject WS (with left recurrent nerve paralysis). Left and right vocal folds are desynchronized which suggests left-right asymmetry of vocal fold properties. Bright regions display glottal opening, dark regions code for vocal fold edge points with glottal closure. The maximum positive excursions are rescaled to unity, the minimum negative excursions are rescaled to zero. The lower part/upper part of each plot corresponds to the temporal evolution of posterior/anterior vocal fold edge points. During four oscillation maxima on the right vocal fold five oscillation maxima appear on the left side. Vertical homogeneity of the vertical bright stripes during glottal opening indicates no relevant anterior-posterior asymmetry.

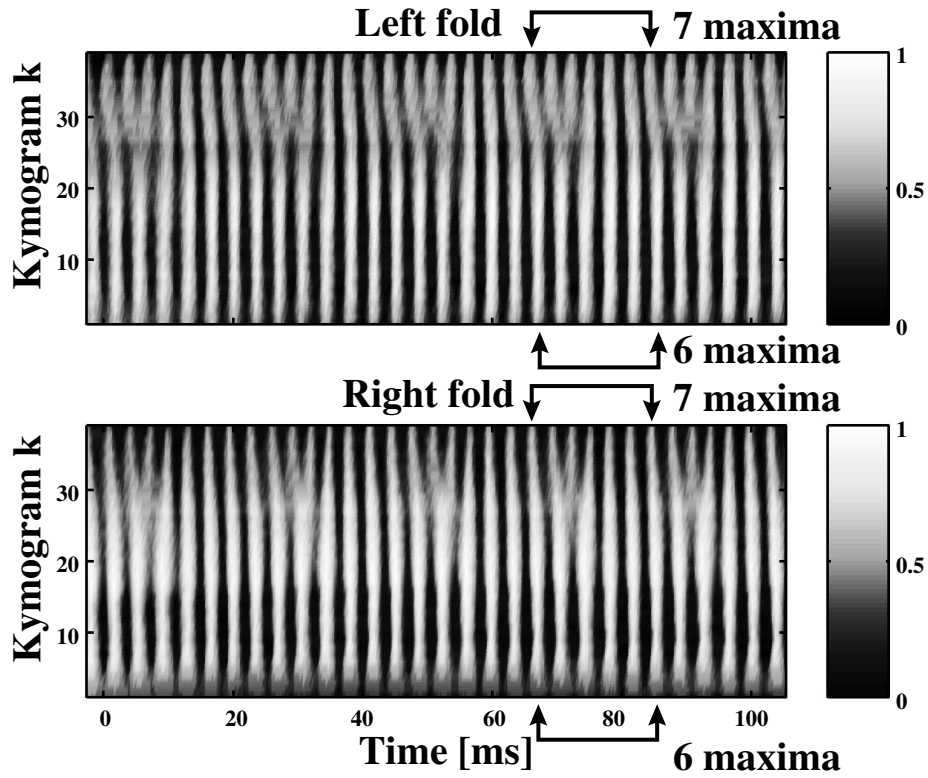


Figure 4.9: Spatio-temporal plots of the scaled elongations $\delta_k^{(\alpha)}(t_i)$ for subject MM (with functional dysphonia). During seven oscillation maxima in the upper part of the spatio-temporal plots, six maxima appear in the lower part. The symmetry between left and right spatio-temporal plot is roughly preserved.

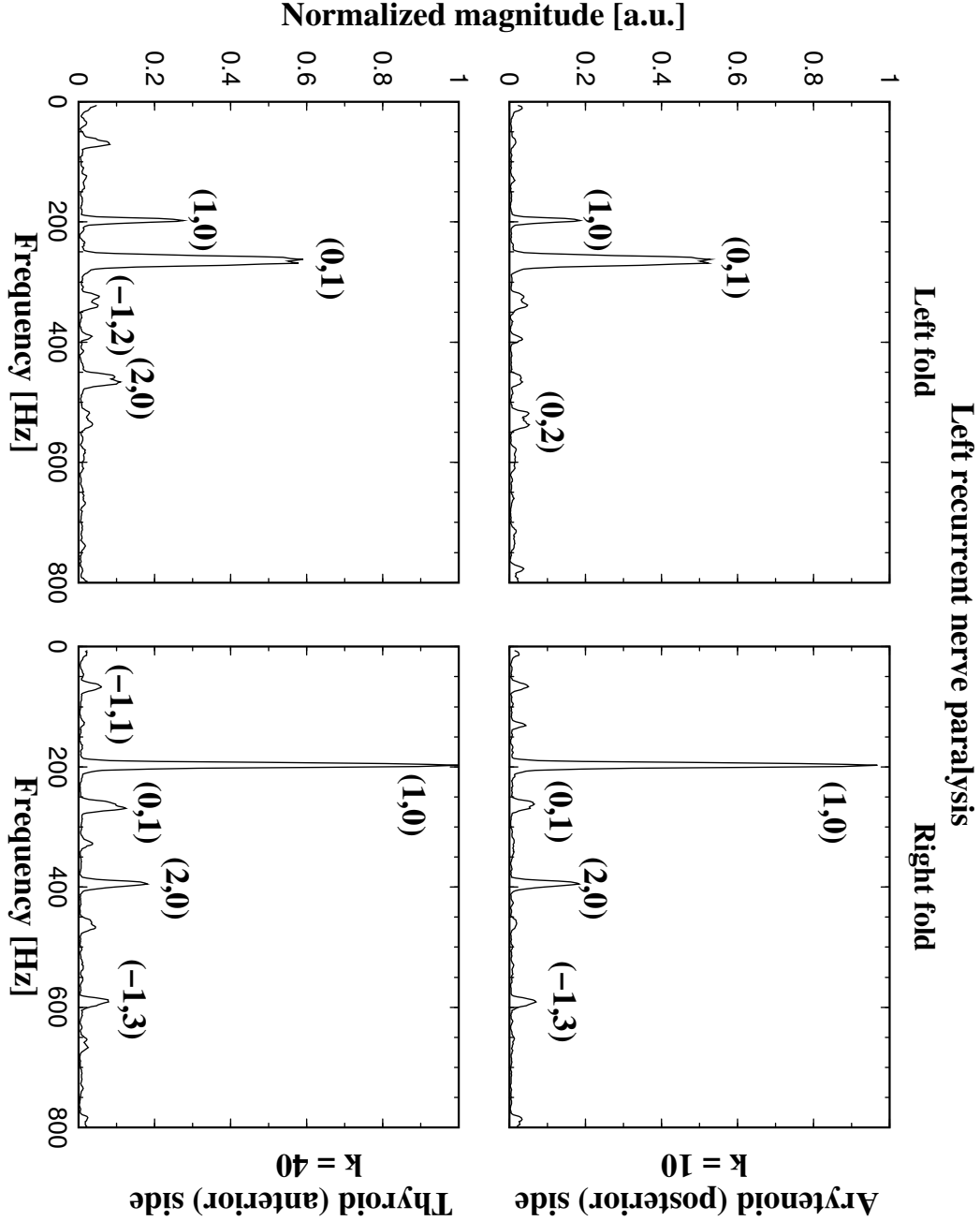


Figure 4.10: Comparison of normalized spectra of $\delta_k^{(\alpha)}(t_i)$ of left and right vocal fold at anterior (kymogram line number $k = 40$) and posterior side (kymogram line number $k = 10$) of each vocal fold (normalization to amplitude maximum of all shown spectra). All amplitude maxima can be explained by linear superposition of two independent frequencies $f_r \approx 197$ Hz and $f_l \approx 265$ Hz : $f_{(m,n)} = mf_r + nf_l$. Left-right asymmetry is $Q_{lr}^{(exp)} \approx 0.74$. The dominant peak frequency for both spectra of the right vocal fold is at f_r , whereas for the left vocal fold both spectra have dominant peak frequencies at f_l .⁷⁶ The spectral peaks are marked by their corresponding numbers (m, n) .

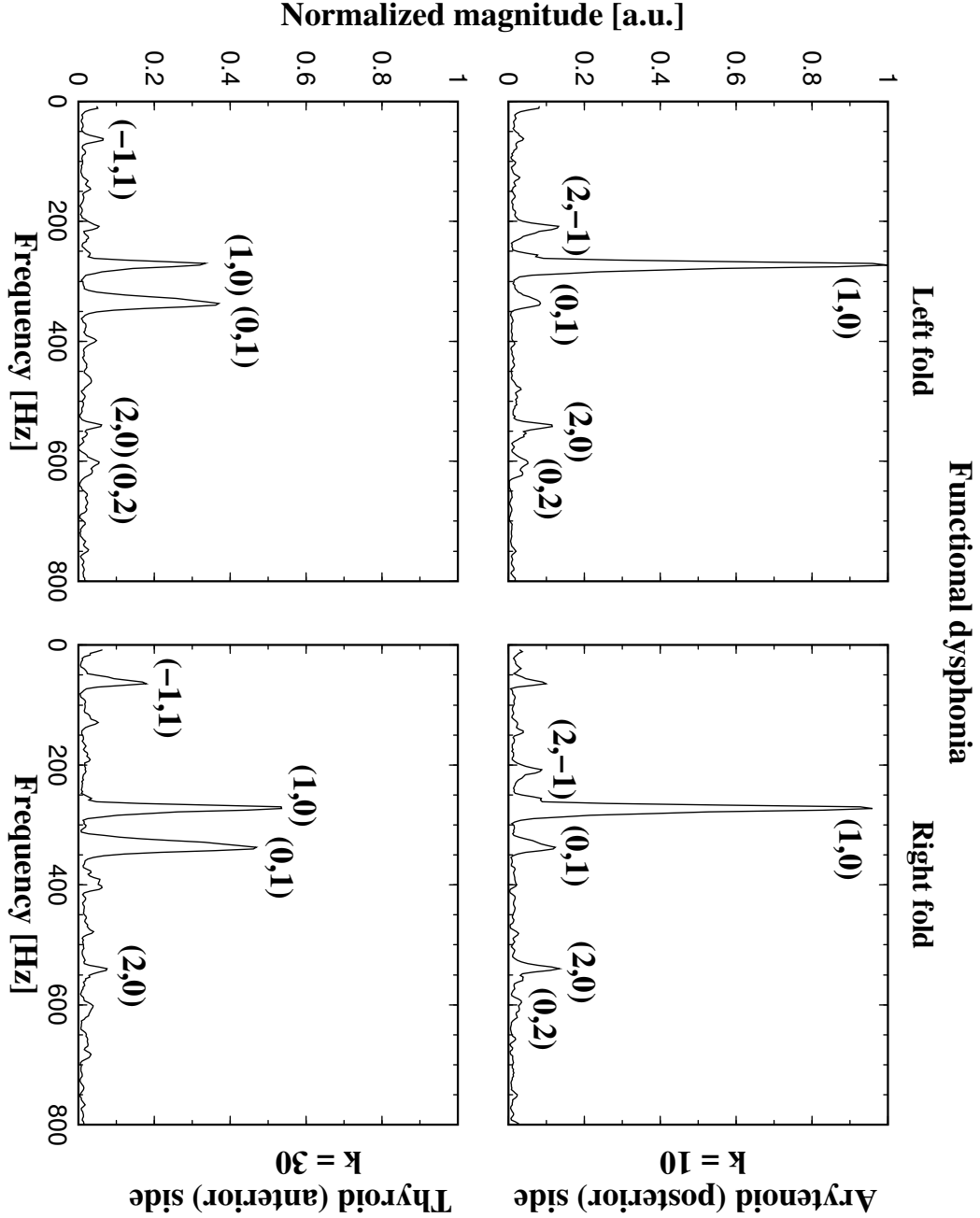


Figure 4.11: Comparison of normalized spectra of displacements $\delta_k^{(\alpha)}(t_i)$ of left and right vocal folds at anterior (kymogram line number $k = 30$) and posterior side (kymogram line number $k = 10$ of each vocal fold (normalization as above): Maxima can be interpreted by linear superposition (m, n) of independent frequencies $f_p \approx 271$ Hz and $f_a \approx 338$ Hz. Anterior-posterior asymmetry is $Q_{ap}^{(exp)} \approx 0.80$.

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

frequency could be found at f_l . On basis of these two chosen time series, anterior-posterior asymmetry was $Q_{ap}^{(exp)} = 1$ indicating no anterior-posterior asymmetry.

In Fig. 4.11 two spectra from subject MM are shown. With these two chosen time series, left-right symmetry was determined to be preserved, $Q_{lr}^{(exp)} = 1$. Moreover, spectra for anterior (kymogram line number $k = 30$) and posterior positions (kymogram line number $k = 10$) differed significantly. Spectra for posterior positions from both left and right side were dominated by the peak frequency $f_p \approx 271$ Hz. At the anterior side, the spectra of both left and right side showed dominant peak frequencies at f_p and $f_a \approx 338$ Hz. Anterior-posterior asymmetry was $Q_{ap}^{(exp)} = f_p/f_a \approx 0.80$. Thus, the desynchronized oscillation of vocal fold edge points within each single vocal fold was observed. Here, the two independent frequencies f_p and f_a were enough to deduce all present peak frequencies.

For the sake of visualization and explanation the number of maxima of the spatio-temporal plots (see Fig. 4.8 and Fig. 4.9) for equal time intervals was determined. The ratio of the number of maxima was different to the ratio of the fundamental frequencies observed by spectral analysis of selected time series of vocal fold edge points. This illustrated the limitation of visual inspection. For example, a superposition of trigonometric functions with a frequency ratio of 4 : 5 could lead to 6 or 8 distinct maxima.

4.3.2 EOF analysis of spatio-temporal vibration patterns

In Fig. 4.12 the relative weights $p_k^{(\alpha)} = \lambda_k^{2(\alpha)} / \sum_l \lambda_l^{2(\alpha)}$ (cf. Appendix A, page 179) of the EOF decomposition are shown for the three phonation examples. The relative weights expressed the relative contribution of different EOFs to the measured spatio-temporal oscillation pattern. The first 21 relative weights were plotted on a logarithmic scale, as they covered about four orders of magnitude. In all three cases the sequence of weights decayed rapidly, whereas higher order weights decreased rather slowly. However, by comparison it could be seen that the decay of the first few weights strongly depended on the phonation example. The cumulative sum of the first five values of the relative EOF weights, given in Tab. 4.2, supported this statement. Thus, the shape of the weight distribution could be used to distinguish the spatial irregularity of different phonation examples. Shannon's entropy $S_{tot}^{(\alpha)}$ (cf. Appendix A, page 179) was used as an overall measure for the shape, i.e. the broadness of the different distributions.

It was found that $S_{tot}^{(\alpha)}$ became significantly larger for the three phonation examples. Correspondingly, the weight distributions became broader. $S_{tot}^{(\alpha)}$ increased from LR and AP symmetrical oscillation (normal phonation) via LR asymmetrical but AP symmetrical oscillation to AP asymmetrical but LR symmetrical oscillation (Fig. 4.12). No significant left-right difference of the spatial irregularities

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

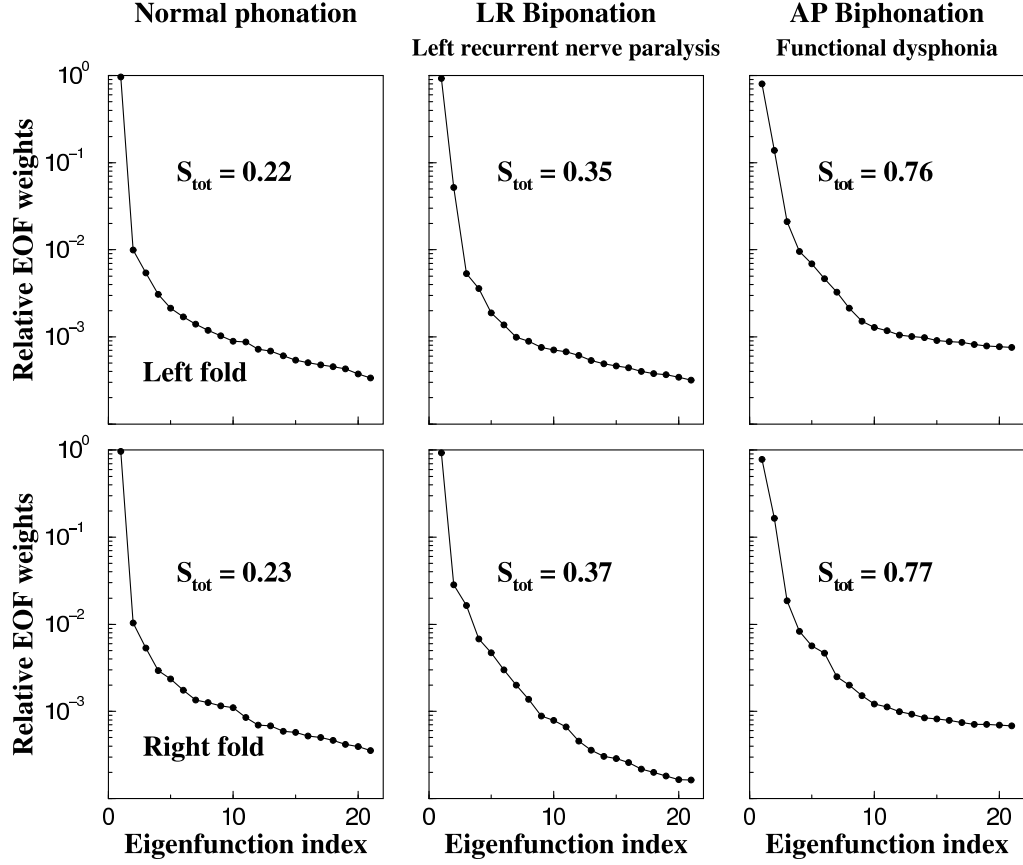


Figure 4.12: First 21 values of the relative EOF weights for three phonation examples for left (upper row) and right (lower row) vocal folds. The information entropy $S_{tot}^{(\alpha)}$ measures the overall spatial irregularity. As explained in the text, the term “LR biphonation” is introduced for the LR asymmetrical phonation of subject WS and “AP biphonation” for AP asymmetrical phonation of subject MM.

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

Eigenfunction index	Cumulative sum $\sum_m \lambda_m^{2(\alpha)}$, $\alpha \in \{left, right\}$ for:					
	LR and AP symmetry (normal phonation)		LR asymmetry (recurrent nerve paralysis)		AP asymmetry (functional dysphonia)	
	left [%]	right [%]	left [%]	right [%]	left [%]	right [%]
1	96.7	96.6	92.8	93.2	80.1	78.2
2	97.7	97.7	97.9	96.1	94.0	94.7
3	98.3	98.2	98.5	97.7	96.1	96.5
4	98.6	98.5	98.8	98.4	97.0	97.3
5	98.8	98.7	99.0	98.9	97.7	97.9

Table 4.2: Cumulative sum $\sum \lambda_m^{2(l/r)}$ for the first five values of the relative EOF weights for three phonation types. For subject JN (with normal phonation) and for subject WS (with LR asymmetrical phonation) the first mode already covers more than 90% of the glottal contour dynamics. However, for the subject MM (with AP asymmetrical phonation) the first mode just carries about 80% of the observed time series. Weights of higher modes become rapidly smaller, but are still specific for the type of phonation. To explain more than 97% of the observed glottal dynamics, two modes have to be taken into account for subject JN, two or three modes, respectively, for subject WS; and four modes for subject MM (indicated by numbers in bold face).

$S_{tot}^{(left)}$ and $S_{tot}^{(right)}$ could be observed.

The number of dynamically relevant modes was estimated choosing a threshold of 97% for reconstruction quality. Tab. 4.2 shows that 97% of the observed time series could be explained by two modes for both vocal folds for subject JN exhibiting LR and AP symmetrical phonation (normal phonation). For subject WS showing LR asymmetrical vibration, it could be found that two modes on the left side and three modes on the right side, respectively, were sufficient. For subject MM exhibiting AP asymmetrical phonation, four modes on both sides had to be included.

Figures 4.13–4.15 display the empirical orthogonal functions associated with the first five weights. For all three phonation examples, the first EOF revealed the uniform outward and inward movement of the vocal folds resulting in a general increase and decrease of glottal opening area. The second EOF illustrated vocal fold displacements with a phase shift of 180 degrees between the anterior and the posterior side of the vocal folds. In most phonation cases their wavelength was

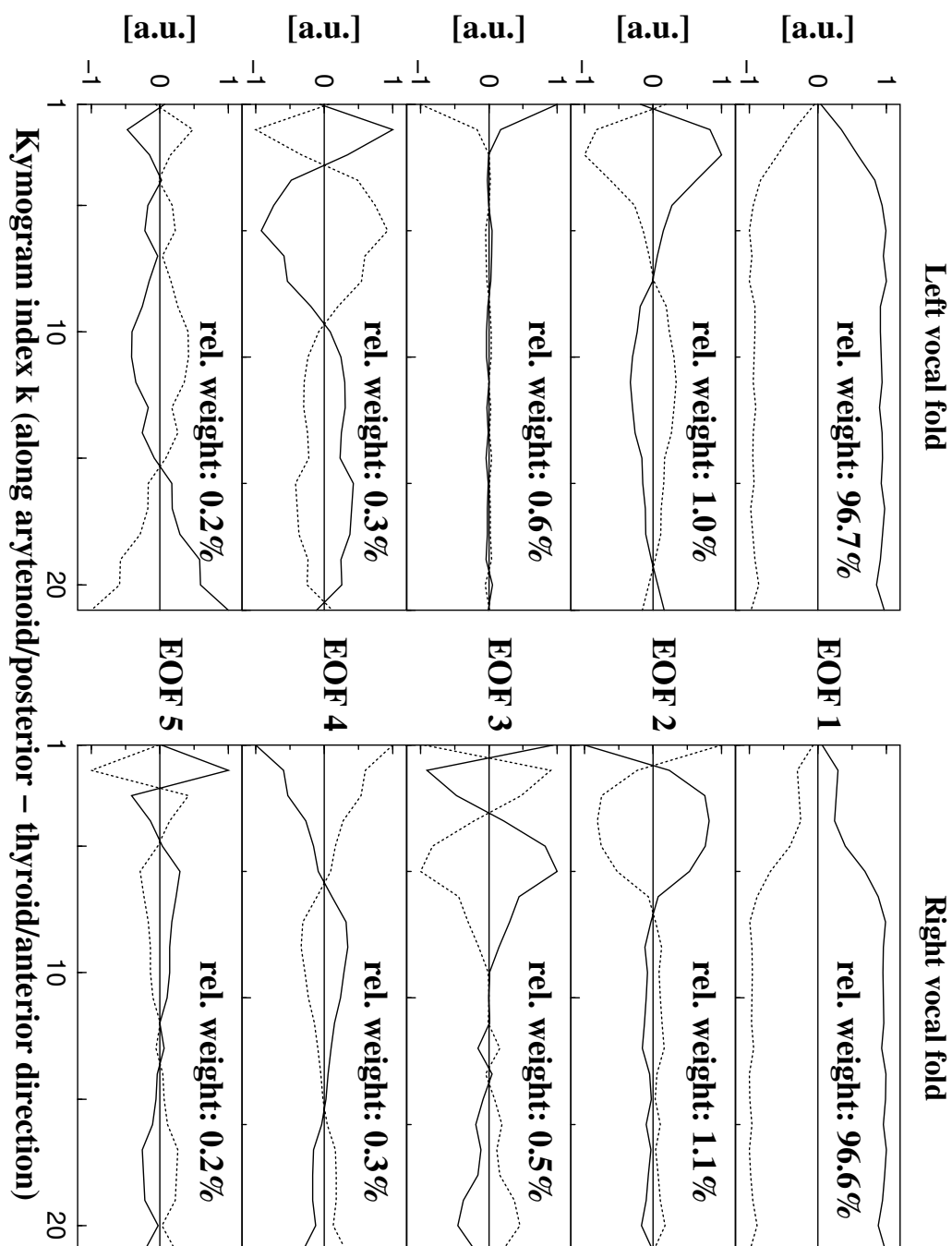


Figure 4.13: First five normalized empirical orthogonal functions for centered time series from subject JN (with normal healthy phonation): Maximum and minimum excursion of the EOFs are shown (normalization to amplitude maxima). The relative weights, indicated in the plots, reflect the contribution of the EOFs to the reconstruction of the observed glottal dynamics. Thus, over 90% of the observed time series is explained by the first EOFs.

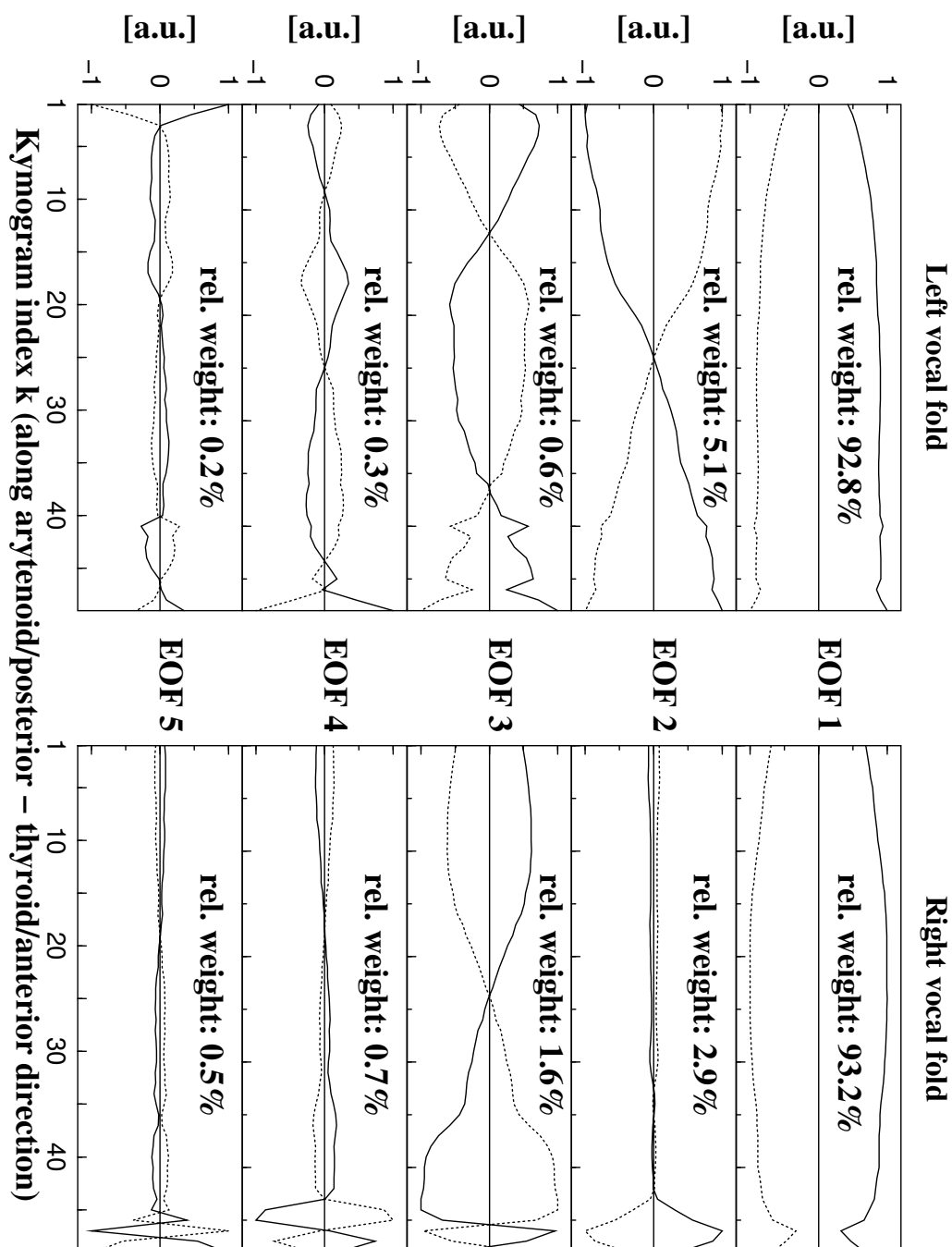


Figure 4.14: First five normalized EOFs for centered time series from subject WS: Maximum and minimum excursion of the EOFs are shown (normalization as above). The first EOFs are enough to explain more than 90% of the time series. Due to scaling the phase shift between anterior and posterior side of the second eof of the right fold can not be seen as clear as on the left fold.

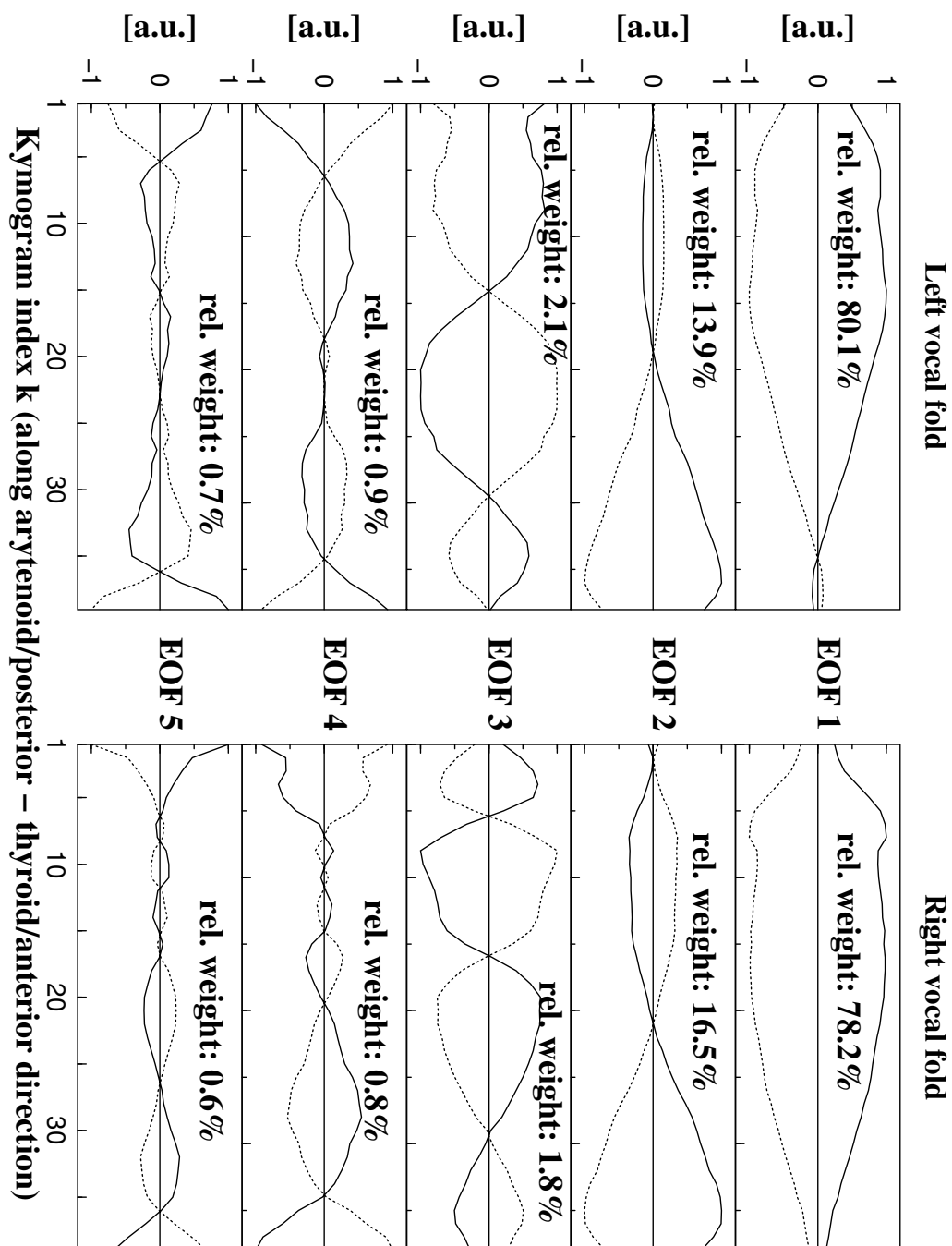


Figure 4.15: First five normalized empirical modes for centered time series from subject MM: Maximum and minimum excursion of the EOFs are shown (normalization as above). The first two EOFs are required to capture more than 90% of the time series.

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

roughly twice the wavelength of the first EOFs. The structure of all higher EOFs was more difficult to generalize due to increasing fluctuations resulting from the finite spatial and temporal resolution of the high speed camera and the noise of the recordings. All modes higher than the first were called “anterior-posterior (AP) modes”.

In Figs 4.16–4.18 the spectra of the corresponding temporal coefficients were plotted. They corresponded to the temporal evolution of the EOFs. For subject JN (regular phonation) it was found, that all peak frequencies of both left and right vocal folds were harmonically related to the fundamental frequency of about $f_0 \approx 127$ Hz. Here, left-right asymmetry $Q_{lr}^{(exp)}$ and anterior-posterior asymmetry $Q_{ap}^{(exp)}$ both were unity. Additionally, it could be seen that temporal coefficients for higher EOFs contained more and more noise.

For subject WS (irregular phonation), the spectra of the first temporal eigenfunction revealed a pronounced left-right asymmetry: The spectrum of the first left temporal coefficient was dominated by the peak frequency $f_l \approx 265$ Hz. The spectrum of the first right temporal eigenfunction had the dominant peak frequency $f_r \approx 197$ Hz. The left-right asymmetry was $Q_{lr}^{(exp)} = f_r/f_l \approx 0.74$. Linear superposition $f_{(m,n)} = mf_r + nf_l$, $m, n \in \mathbb{Z}$ of the two independent left and right frequencies explained the peak frequencies of all remaining spectra apart from noisy contributions.

For subject MM (irregular phonation), the first two spectra of the temporal eigenfunctions were roughly symmetric about left and right side. On each vocal fold side, however, a significant difference in peak frequency for the first two temporal coefficients could be observed. For both first temporal coefficients, the peak frequency was $f_p \approx 271$ Hz. Both spectra of the second temporal coefficients had a peak frequency $f_a \approx 338$ Hz. The indices “p” and “a” for f_p and f_a were used, because the main spatial contribution of the first (and second) EOF was observed in the posterior (and anterior) part of the vocal folds (see Fig. 4.15). The anterior-posterior asymmetry was $Q_{ap}^{(exp)} = f_p/f_a \approx 0.80$. Linear combinations $f_{(m',n')} = m'f_p + n'f_a$, $m', n' \in \mathbb{Z}$ of these independent frequencies explained all remaining peak frequencies in the shown spectra.

4.4 Discussion

The aim of this study was to analyze spatially complex and temporally irregular glottal contour patterns of pathological phonation. Mode analysis of spatio-temporal patterns of vocal fold edge displacements *in vivo* with empirical orthogonal functions appeared to be an appropriate tool to extract principal glottal vibration modes from high-speed recordings. EOF analysis can extract excited modes from measured time series. Although this method is applied to analyze only three

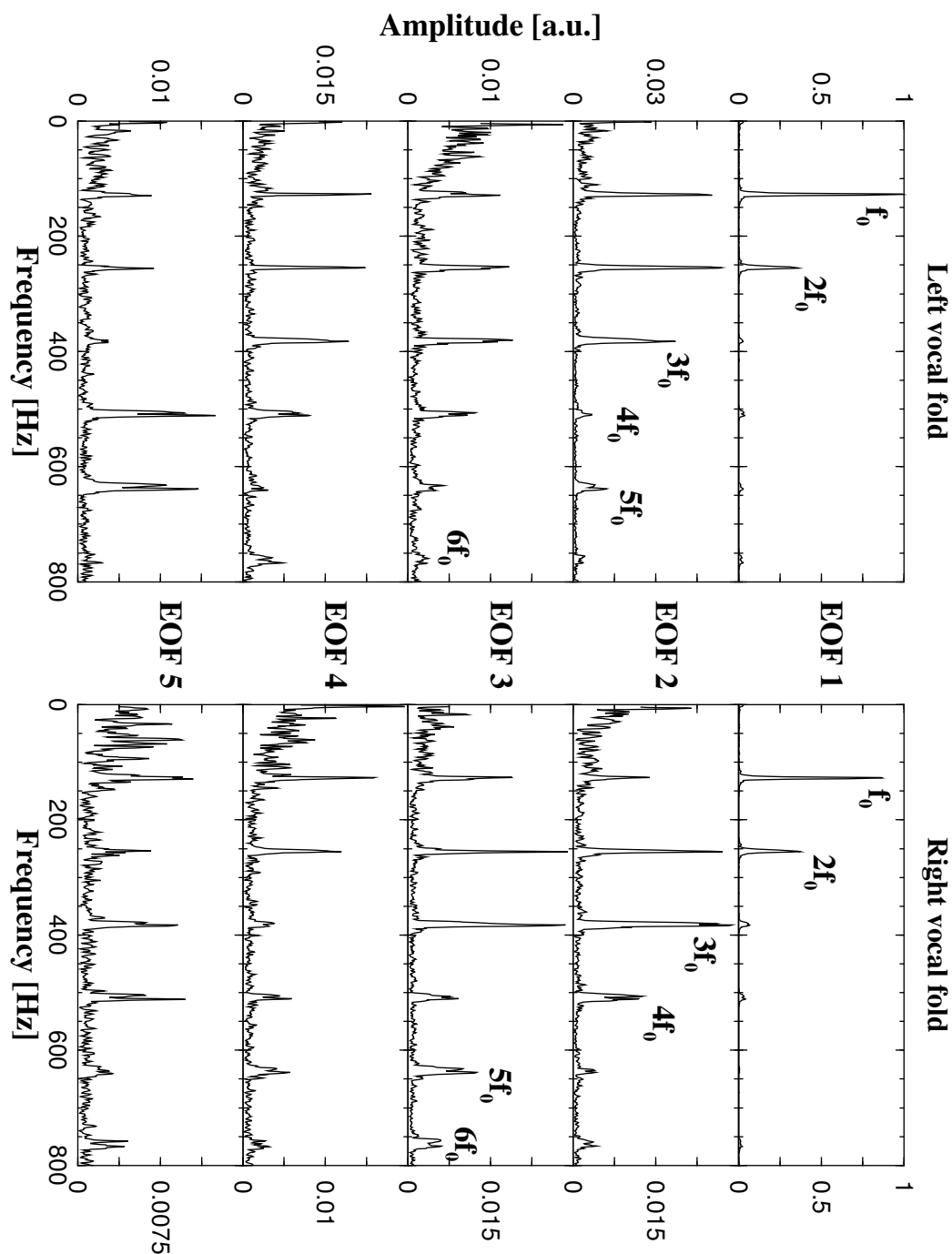


Figure 4.16: Normalized magnitude spectra (linear scale) of first five temporal eigenfunctions for subject JN (with normal healthy phonation) (normalization to amplitude maximum of all shown spectra): All spectra consist of harmonically related peak frequencies together with noise contributions increasing with EOF number.

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

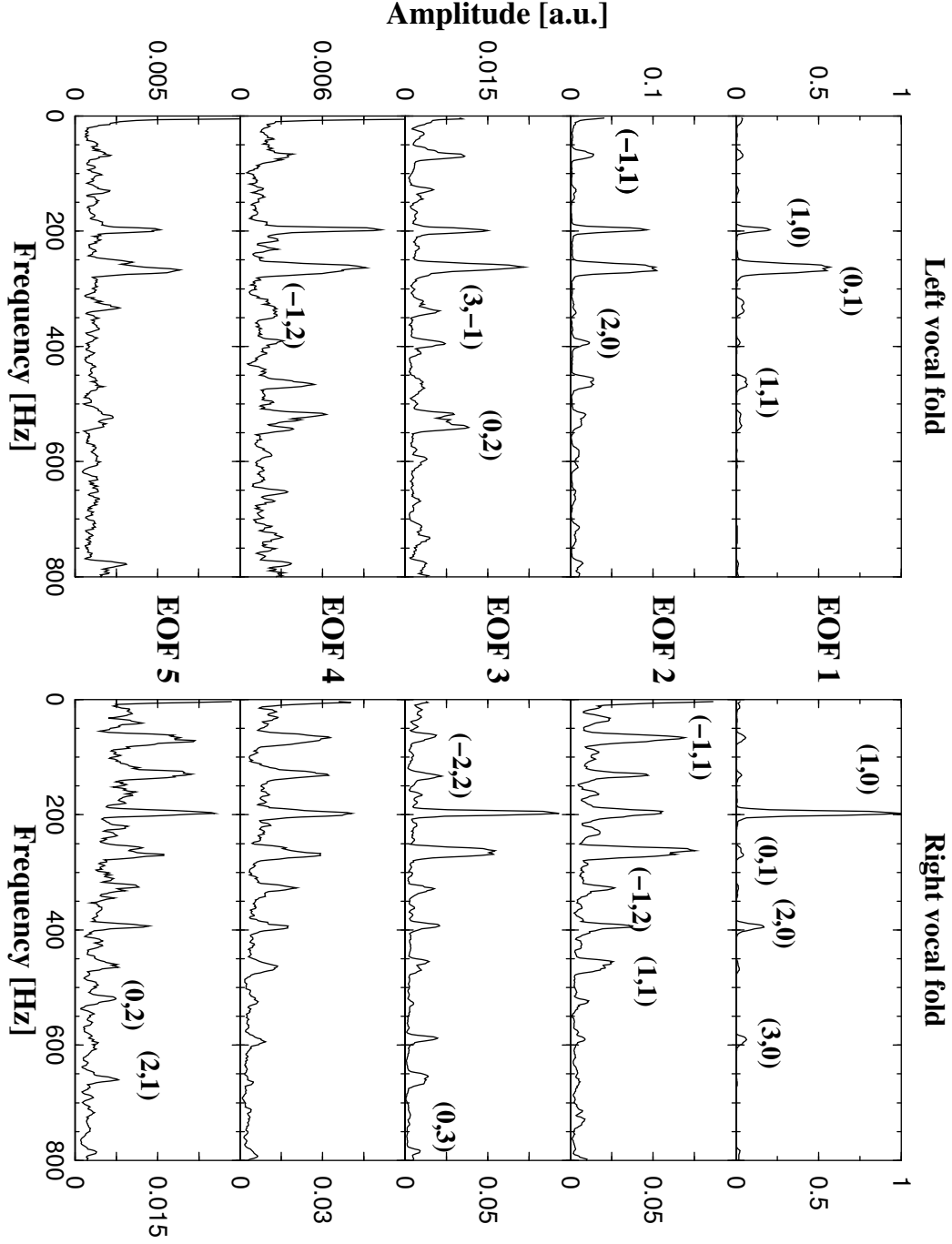


Figure 4.17: Normalized magnitude spectra of first five temporal eigenfunctions for subject WS (normalization as above): The left-right asymmetry ratio $Q = f_r/f_l = 0.74 \approx 4/5$ observed in the first temporal eigenfunctions associated with the first EOF reflects the laryngeal asymmetry. The main spectral information is already contained in the first temporal eigenfunctions.

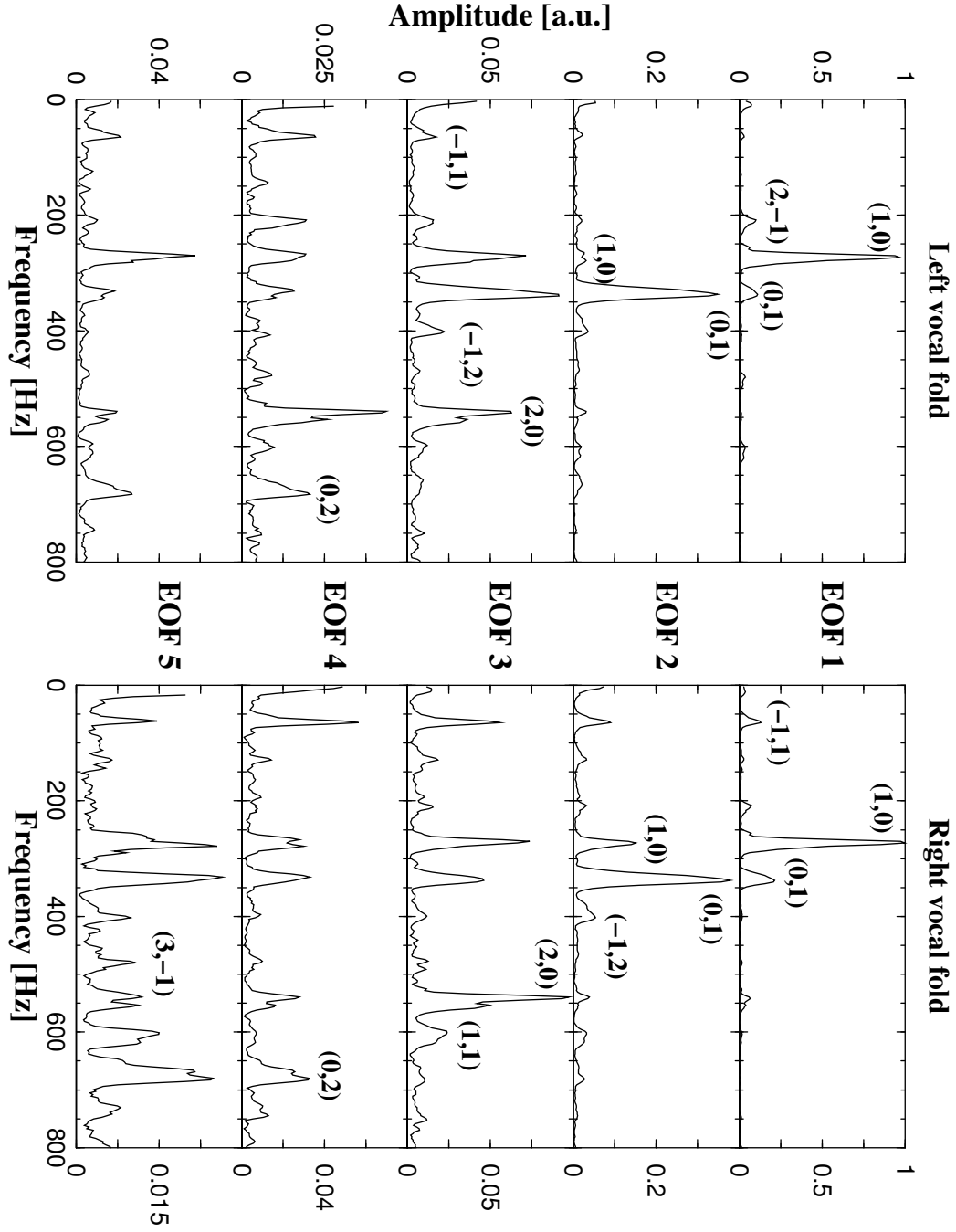


Figure 4.18: Normalized magnitude spectra of first five temporal eigenfunctions for subject MM (normalization as above): The left-right asymmetry ratio $Q = 1$ of the first two temporal eigenfunctions indicates laryngeal left-right symmetry. AP asymmetry is indicated by the qualitative difference in the spectral content of the first and the second temporal eigenfunctions. Therefore, the main spectral information is contained in the first two EOFs.

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

Subject	Left-right asymmetry	Anterior- posterior asymmetry	Spatial irregularity		Number of relevant modes	
	$Q_{lr}^{(exp)}$	$Q_{ap}^{(exp)}$	$S_{tot}^{(l)}$	$S_{tot}^{(r)}$	left	right
JN	1.0	1.0	0.22	0.23	2	2
WS	0.74	1.0	0.35	0.37	2	3
MM	1.0	0.80	0.76	0.77	4	4

Table 4.3: Summary of measures quantifying spatio-temporal vocal fold vibration patterns for three phonation examples: Phonation patterns are described by the left-right asymmetry, anterior-posterior asymmetry, the overall spatial irregularity, and the number of dynamically relevant EOF modes. The number of relevant modes is related to the spatial irregularity. Thus it could be connected to the minimum number of degrees of freedom a prospective biomechanical model for the individual pathology should provide.

examples of normal and pathological phonation, the presented method and the shown effect of left-right desynchronization and anterior-posterior desynchronization are of general validity. Previous theoretical studies of biomechanical models (Titze et al., 1993; Herzel, 1993) support this point of view.

This chapter presents the analysis of two different types of biphonation generated by different glottal mechanisms. Biphonation is already known to be induced by left-right asymmetry of the vocal folds, and strong interaction with the supraglottal vocal tract is assumed to ease biphonation (Titze and Story, 1997; Mergell and Herzel, 1997b). Data from a patient with left recurrent nerve paralysis is presented, where the desynchronization of left and right vocal fold oscillations was observed. It is found that biphonation, i.e. the existence of two fundamental frequencies, can be explained in this case by only the first mode of the EOF decomposition of left and right oscillations. This result strongly supports the notion that biphonation was induced by a left-right asymmetry of the vocal folds due to the paralyzed left vocal fold. This example was presented to contrast it with the finding of another mechanism for biphonation.

Therefore, data on a patient with functional dysphonia is presented, where desynchronized oscillations of vocal fold edge points on each vocal fold were observed. In addition to the uniform inward and outward movement of the vocal folds, oscillation patterns with a prominent phase shift between anterior and posterior side of the vocal folds were observed. It was found that these higher modes vibrate independently from the basic spatially uniform mode as two independent

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

oscillation frequencies can be seen in the spectra of the temporal coefficients of the EOFs. Thus, experimental evidence is provided that desynchronization of modes can be observed in voice disorders, as previously suggested by theoretical considerations (Titze et al., 1993; Herzel et al., 1994). Theoretically, the vocal folds can be considered as complex, visco-elastic, three-dimensional structures, that can generate complex spatio-temporal vibration patterns by many different oscillating structures. Biphonation, so far experimentally known to result from the desynchronized oscillation of the entire left and right vocal folds (Mergell et al., 2000), is extended to toroidal behavior of two oscillating structures on single vocal folds. For a clearer distinction, the term “LR biphonation” is introduced for the first case, and the term “AP biphonation” for the second case to express the interaction of the first EOF with different AP modes (i.e. higher EOFs) that generates biphonation.

Four measures were provided, that characterize spatio-temporal oscillation patterns observed in digital high-speed recordings: the left-right asymmetry Q_{lr} , the anterior-posterior asymmetry Q_{ap} , the information entropy S_{tot} as spatial irregularity, and the number of relevant modes. Table 4.3 summarizes the results for the shown phonation examples. For normal phonation, both Q_{lr} and Q_{ap} are equal to unity, and the entropy and the number of modes are small. For LR asymmetrical vibration, Q_{lr} is significantly smaller than unity (due to the higher fundamental frequency of the paralyzed vocal fold). Within measuring accuracy, Q_{ap} is unity. The spatial irregularity, described by S_{tot} , and the number of modes is significantly larger than for normal phonation. It is concluded, that due to the spatial irregularity higher modes are excited, which are synchronized within single vocal folds, but desynchronized between either vocal folds (due to the paralysis of the left vocal fold). For the AP asymmetrical oscillation, Q_{ap} is significantly smaller than unity, and Q_{lr} is unity. The increased entropy S_{tot} and the larger number of relevant modes, compared to recurrent nerve paralysis, indicate a further increase in spatial irregularity. It is concluded that on either vocal fold higher modes are excited. These modes can vibrate independently, presumably due to invisible anterior-posterior inhomogeneities. This study suggests that modes on both vocal folds could be synchronized due to short moments of vocal fold contact and the airstream. This conclusion deserves further investigation analyzing more pathological phonation examples showing anterior-posterior biphonation.

In this study methods are described to analyze irregular spatio-temporal vocal fold oscillations quantitatively. First spectral analysis was applied to time series of selected vocal fold edge points along the anterior-posterior direction. In general this method is not appropriate to decompose spatio-temporal oscillation patterns as the choice of the time series is quite arbitrary and the interdependence between different time series associated with different edge points along the anterior-posterior direction is lost. However, this approach can provide first hints

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

to the underlying mechanism of irregular vocal fold oscillations. Nevertheless, for a quantitative and consistent spatio-temporal analysis we used empirical orthogonal functions to decompose the HGG data into different modes. This concept incorporates correlations of different time series and resolves the dynamics of different spatial modes. The results for the left-right asymmetry and the anterior-posterior asymmetry with both methods were consistent. Nevertheless, spatial irregularity could be quantified only using empirical orthogonal functions to decompose the observed spatio-temporal patterns into a hierarchy of different spatial modes. Of course, this method works similar to a rough and just qualitative analysis of subsequent high-speed frames done by an observer with less sophisticated methods.

Data on three subjects were shown with normal phonation, LR biphonation and AP biphonation. The terms “healthy phonation”, “left recurrent nerve paralysis” and “functional dysphonia” were used just as labels to discriminate between the time series. It is beyond the scope of this study to discuss the connection of the pathophysiology of the subjects with the observed phonation patterns. Nevertheless, the presented analysis method may help to replace the ill-defined term “functional” by a more detailed diagnosis. In contrast to cases with clearly visible morphological changes of the vocal folds the term “functional dysphonia” is used for pathological phonation with no apparent structural change in the larynx (Kotby et al., 1993; Wittenberg et al., 1997). So until now “functional” just states the absence of diagnostic methods to describe the pathology on basis of direct observations. This study suggests to use the EOF analysis to quantify the dynamics of hoarseness related to dysfunctions of the laryngeal configuration hidden in the complex vibratory patterns.

As a possible mechanism for the desynchronization of the first two modes, it can be suggested that inhomogeneities of the vocal fold tissue properties, such as local hidden morphological changes in deeper tissue layers could induce and support the independent vibration of modes. These endoscopically invisible changes in vocal fold tissue properties might decrease the coupling between anterior and posterior parts of the vocal folds facilitating biphonation. Here, the term inhomogeneities is used to express the deviance (like local morphological changes) from the normal inhomogeneity occurring along the anterior-posterior direction together with the different layer structure.

To a certain extent, the recorded HGG time series are contaminated with noise originating from different sources. Examples are finite temporal and spatial sampling accuracy, and thermal noise of the CCD chip used in the digital camera. Uncertainties of the vocal fold edge detection process due to blurred recordings and to mucus or light tissue structure on the surface of the vocal folds are further noise sources. Noisy contributions are found in the EOF modes, in their weights and in their temporal coefficients. Due to noise, the sequence of weights of higher

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

modes of the EOF decomposition decrease rather slowly (North et al., 1982; Landa and Rosenblum, 1991). As the generation of noise is not modeled, a simple and robust threshold criterion is used to estimate the dynamically relevant modes. Thus, the number of relevant modes of different phonation examples and of previous theoretical work can be compared. For a more sophisticated criterion, more information about mechanisms generating noise in the digital high-speed observations have to be considered (see Landa and Rosenblum, 1991). As the weights of higher modes have negligible values due to noise, only the first 21 weights were used to calculate Shannon's entropy S_{tot} and to plot the weight distribution (Fig. 4.12) for the sake of comparability.

Using endoscopy and a high-speed setup several systematic errors of the recording procedure and the extraction of time series of vocal fold edge points have to be taken into account. For a detailed discussion of blurring due to finite temporal and spatial resolution, technical problems with rotation, vertical and horizontal relative movement between glottis and endoscope, and the ambiguity of vocal fold edge detection due to inferior-superior vibratory mode see previous work (Wittenberg et al., 1997; Wittenberg, 1998). Regarding the finite temporal and spatial resolution the analysis must be restricted to a range given by Shannon's sampling theorem. Furthermore, the influence of higher frequency components is neglected because they cannot be resolved by the finite spatial and temporal resolution of the high-speed camera system. However, since EOF analysis is based on averaging, random fluctuations are suppressed.

EOF analysis of time series obtained from three-dimensional biomechanical simulations of normal and chaotic vocal fold oscillations (Berry et al., 1994) show that two modes are enough to capture the vibration pattern of normal phonation. In these extensive computer simulations of a model with 414 degrees of freedom, four dominant EOFs could still explain even more complex dynamics. Similar results were found analyzing *in vivo* data. As HGG data of only the upper vocal fold edge is analyzed, the first two modes found in the biomechanical model (Berry et al., 1994) nearly correspond to the first EOF of the *in vivo* data. The major limitation of *in vivo* data obtained with the high-speed setup is that time series from the endoscopic view show the dynamics of mainly the upper vocal fold edge. Therefore, all modes related to the vertical vocal fold direction cannot be distinguished.

This work proposes a method for a systematic development of appropriate low-dimensional biomechanical models. These models could incorporate left-right asymmetries and anterior-posterior inhomogeneities to simulate irregular spatio-temporal vibration patterns. Multiple trajectories of vocal fold edge points could be reduced to a few principal vibrating modes, which could then be used in model simulations. The number of degrees of freedom of appropriate low-dimensional

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

models should correspond to the number of principal vibrating modes observed *in vivo*. Here, for subject JN the measured dynamics from 21 vocal fold edge points on each vocal fold can be reduced to the dynamics of two modes. Similarly, the observed dynamics of 49 glottal edge points of subject WS could be compressed into two or three modes respectively. For subject MM, a reduction of the dynamics of 39 observed edge points to four principal modes could be obtained. Thus, models simulating AP mode dynamics might be derived in a systematic way from high-speed observations, where inhomogeneities of the visco-elastic properties of the vocal folds, e.g. localized morphological changes, could be incorporated. This hypothesis deserves further investigation.

The clinical motivation for using EOF analysis was to measure “AP modes” known from literature and frequently observed qualitatively in clinical everyday work (Hess et al., 1994; Tigges et al., 1999). The main impact of EOF analysis is the measurement of the dynamics of spatial modes crucial for the description of ap modes. Thus, empirical orthogonal analysis could reveal endoscopically invisible morphological and, moreover, functional changes of myoelastic properties of the vocal folds. Moreover, it could be expected that in other cases also visible modifications such as cicatrices from operations, small polyps, nodules, cysts or tumors may effect nodes and antinodes of surface waves in the surface tissue of the vocal folds. Local stiffening of visible superficial vocal fold tissue or local stiffening or atrophy of invisible deeper mucosal and muscular tissue may also constrain certain modes of vocal fold vibration. With EOF analysis, aerodynamical and visco-elastical excitation mechanisms for AP modes may be examined. The coupling of AP modes and the sub- and supraglottal tract can be studied as well. Thus, with EOF analysis a more detailed diagnosis for pathological phonation may be provided using the measures proposed in this study. Moreover, EOF analysis may give hints for surgery and speech therapy. From the viewpoint of nonlinear dynamics, these suggestions merit a closer analysis of models that exhibit AP modes. Excitation mechanisms and the bifurcation structure of vocal fold models revealing spatio-temporal vibration patterns are of clinical importance for the diagnosis of irregular phonation.

4.5 Conclusion

This study shows that spatio-temporal glottal contour patterns obtained from high-speed observation of pathological phonation can be decomposed into modes using EOF analysis. Therefore, laryngeal asymmetries can be quantified. Previous methods, that analyzed time series of single vocal fold edge points (Mergell et al., 2000), have been extended to analyze the complete vocal fold contour, i.e. the entirety of the upper vocal fold edge points. Also the technique of multi-line

4. Spatio-temporal Analysis of Irregular Vocal Fold Oscillations

kymography described by Tigges et al. (1999) has been extended.

A new generation mechanism for the known phenomenon of biphonation has been found. This study has shown experimentally that so called AP modes are excited during vocal fold oscillations, and the weights of AP modes are quantified. Here it has been observed that AP modes oscillate independently from the basic glottal mode. This behavior leads to biphonation. Thus, experimental evidence is provided for the desynchronization of modes previously observed in theoretical models (Titze et al., 1993; Berry et al., 1994; Alipour-Haghighi et al., 2000) and in *in vitro* experiments, e.g., with excised larynges (Berry, 2001). Thus, the presented method of examining high-speed observations of vocal fold vibrations *in vivo* complements previous findings in computer models and *in vitro* high-speed studies in the laboratory. This study suggest the use of EOF analysis to systematically develop simple low-dimensional biomechanical models of vocal folds using direct endoscopic high-speed observations of vocal fold oscillations *in vivo*.

Chapter 5

Modeling the Dynamics of Vocal Membranes in Bats and Primates

Vocal membranes are one widespread morphological variation of vocal folds in nonhuman mammals. They are thin lightweight upward extensions of the membranous portion of the vocal folds. In bats they produce ultrasonic echolocation calls. In nonhuman primates vocal membranes are responsible for the highly diverse call repertoire. In bats and primates complex vocalizations, such as subharmonic oscillations, biphonation, irregular chaotic calls, register jumps, pulsed high frequency oscillations, and abrupt transitions between these different behaviors are frequent. In this study a phenomenological vocal membrane model was developed to understand the production of these complex vocalizations. The well-known simplified two-mass model was extended by oscillating vocal membrane plates attached to the upper edge of the vocal folds. Linear eigenmode analysis revealed the influence of the vocal membrane geometry on eigenfrequencies and eigenmodes. These results were interpreted on the background of the theory of coupled oscillators. Parameter regions with frequency locking were identified. The influence of vocal membrane geometry on phonation onset was discussed based on the eigenmodes of the model. Hopf bifurcation analysis revealed phonation onset pressure and onset frequency. Two different voice registers were found. Overlapping register areas indicated biphonation (toroidal oscillation) in the vocal membrane model. Numerical integration with slowly varying model parameters supported this finding. Hopf bifurcation analysis showed that vocal membrane design could minimize phonation onset pressure and enlarge phonatory pressure ranges. Numerical simulations with gliding subglottal pressure variation and varying frequency tuning (related to the ratio of oscillation frequencies of vocal

folds and the vocal membrane) showed voice instabilities that qualitatively resembled observed vocalization patterns in bats and primates.

5.1 Introduction

Sound production and speech information coding in humans have been extensively studied in the last few decades. The principles of normal voice and speech production have been widely understood (Titze, 1994a). Moreover, pathological voice production has been described both qualitatively and quantitatively (Ishizaka and Isshiki, 1976; Behrman and Baken, 1997; Giovanni et al., 1999a; Mergell et al., 2000; Neubauer et al., 2001).

In the previous two chapters of this dissertation, two aspects of phonation have been studied: the classification of the artistic use of the nonlinear phonatory system and nonlinear phenomena in voice pathologies resulting in complex spatio-temporal oscillation patterns. This chapter focuses on modeling voice instabilities in nonhuman mammalian phonation.

Influenced and propelled by advances in human voice research, the biomechanical principles of animal vocalization and communication systems have gained more and more interest. Central to animal communication, nonhuman phonation and sound production have been studied both experimentally and theoretically (oscine birds: Fletcher (1988); Fee et al. (1998); Gardner et al. (2001); Laje et al. (2002); primates and bats: Mergell et al. (1999)).

Although voice production in animals is central for animal communication, there are very few studies on nonhuman sound production and phonation. In nonhuman mammals, there is one widespread morphological variation of the vocal fold anatomy – vocal membranes. Thus far, a theoretical study of the dynamics of vocal membranes is missing. This chapter aims at a quantitative understanding and qualitative reproduction of vocalization patterns in nonhuman mammals with vocal membranes.

5.1.1 Significance of a biomechanical vocal membrane model

Modeling vocal membranes in nonhuman mammals bridges fields of biology, neuroscience and physics. It includes data fitting and addresses evolutionary and functional questions. In general, the adaptability and robustness of a communication system should be explored. Which sounds are easy, or impossible, to produce for nonhuman mammals? What are the costs and benefits of a given sound type in terms of energetics, predator detection, environmental transmission, and receiver characteristics? To date theoretical studies of the dynamics of vocal membranes are missing. Theoretical modeling addresses questions about the evolutionary signifi-

5. Vocal Membrane Model for Bats and Primates

cance of vocal membranes in nonhuman mammals. Why do nonhuman mammals need vocal membranes? Why did they not disappear during evolution? Is there an advantage or a selective pressure for the conservation of vocal membranes? Regarding typical communication patterns in nonhuman mammals, what is the role of vocal membranes in producing these patterns?

In this chapter basic signal production mechanisms of a simplified model for vocal membranes are studied. This will eventually give insights into the evolution of acoustic communication in nonhuman mammals. In this chapter a vocal membrane model will be developed and analyzed. It will be shown that this model is able to reproduce vocalization patterns, such as pulsed high frequency oscillations, voice registers, subharmonic oscillations, biphonation and chaotic oscillations, frequently found in nonhuman mammals with vocal membranes.

5.1.2 Vocal membranes in nonhuman mammalian larynges

Constrained by the functional requirements of airway protection and respiration, the gross anatomy and physiology of the mammalian larynx shows little qualitative variation from species to species (Negus, 1949; Harrison, 1995; Schön-Ybarra, 1995; Fitch and Hauser, 1995). Compared with the human larynx, there are two main morphological changes in nonhuman mammalian larynges: air sacs and vocal membranes. A description of the morphological diversity and a discussion of functional relevance of air sacs can be found, e.g., in Negus (1949); Gautier (1971); Fitch and Hauser (1995).

Vocal membranes (also termed “vocal lips”) are a widespread variation of vocal fold morphology in nonhuman mammals. They are thin lightweight upward extensions of the membranous portion of the vocal folds. They consist of connective tissue without muscle fibers. Most microchiropteran bats have vocal membranes (see Fig. 5.1) that can be as thin as a few microns across (Suthers and Fattu, 1973; Suthers, 1988). Primates are another large mammalian group with vocal membranes, for example, most New World (platyrrhine) monkeys, some Old World (catarrhine) monkeys, and the apes (Griffin, 1958; Schön-Ybarra, 1995; Starck and Schneider, 1960). Furthermore, vocal membranes are known to be present in members of the genus *Felis* (smaller members of the cat family (Hast, 1971)), and in llamas and young pigs. Upward extensions of vocal folds resembling vocal membranes have been reported in dog-wolf mixes (Riede et al., 2000b), hyaenas, and canids (Riede et al., 2000a).

5.1.3 Vocal membranes in bats

This study was motivated by anatomical and acoustical data from microchiropteran bats. Many bats use ultrasonic echolocation (produced by vocal membranes) to

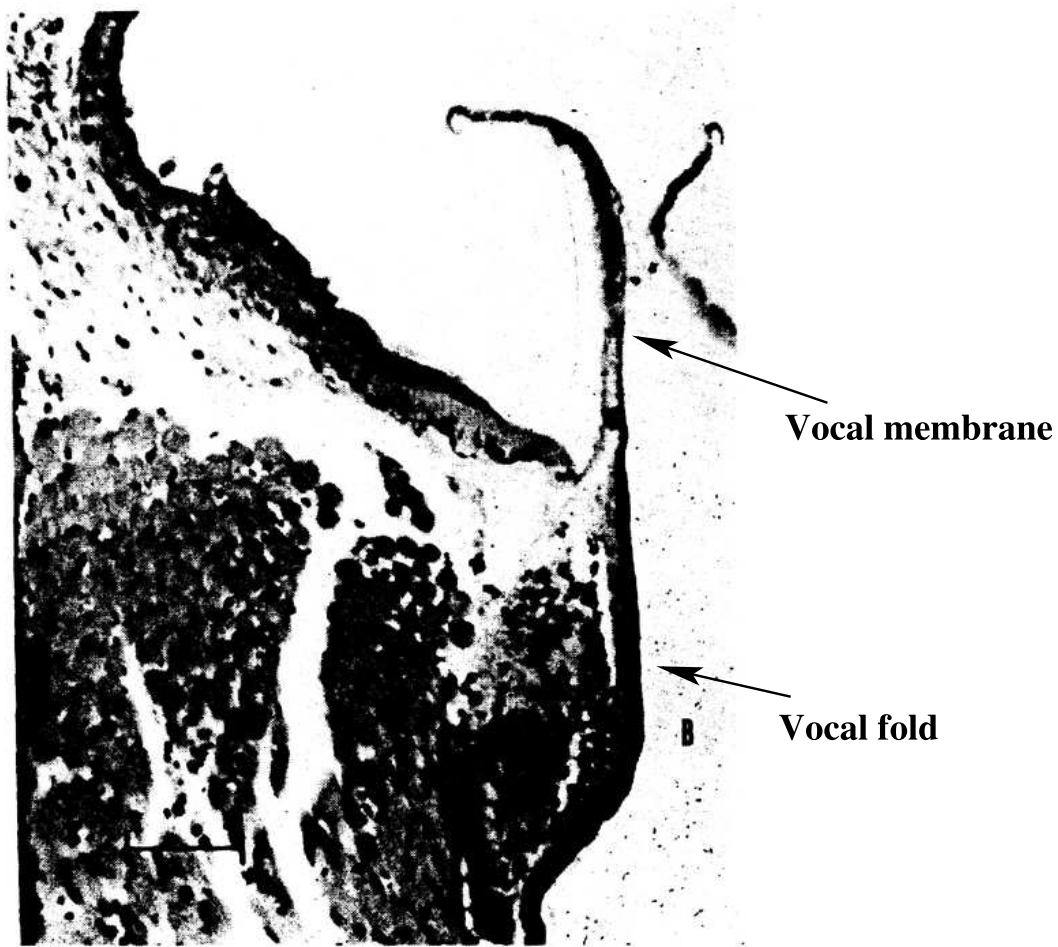


Figure 5.1: Cross section of vocal fold with attached vocal membrane of bat *Eptesicus fuscus*, bar equals 50 μm (from Suthers and Fattu, 1973)

detect prey and obstacles. For these tasks they use calls of fundamental frequencies between 11 and 212 kHz (Jones, 1999).

The sound repertoire and behavioral context of echolocating bats is very well documented. A broad database of spectrographic displays of bat vocalizations is readily available. Although the basic mechanisms for ultrasound production are understood (Griffin et al., 1960), a detailed modeling of sound production is still missing.

Echolocating bats using pulsed sounds are able to change the pulse rate over a wide range from a few Hz up to 200 Hz. During terminal feeding buzzes of insect catching bats, the typical fundamental frequency of downward frequency

5. Vocal Membrane Model for Bats and Primates

sweeps can sharply drop accompanied by a sharp increase of repetition rate up to 200 Hz (Britton and Jones, 1999; Schnitzler et al., 1987; Hartley et al., 1989). It is unclear whether high pulse repetition rates are due to fast neural control of each single pulse (discussion in Griffin et al. (1960)), or whether they are an intrinsic dynamic feature of the primary sound production apparatus. The first hypothesis is supported by tetanus frequency measurements in *in vivo* muscle fibers (Griffin et al., 1960; Suthers and Fattu, 1973). The alternative hypothesis could be a dynamic mechanism involving wavelike movement of the vocal folds at low pulse frequencies that gates the high ultrasound frequency oscillation of vocal membranes. In the latter case the observed frequency sweeps could be produced by intrinsic dependencies of stiffness, oscillating masses, and effective vocal membrane height on the phase of the oscillating and gating vocal folds. Unfortunately, direct observations of pulsed echolocation calls with high speed cameras are still missing.

Echolocation, with respect to *in vivo* and *in vitro* sound production and perception, has been extensively studied in several species (e.g., Erwin et al., 2001). Many publications focus on the descriptive statistics of ultrasonic pulse emission, center pulse frequencies, minimum and maximum frequency in frequency sweeps and pulse rates during prey capture in bat species ranging from insect catching bats to plant and nectar feeding bats. An extended list of topics related to this modeling study along with a more detailed review of significant papers can be found in Appendix B.1, page 182.

Important findings from these studies (Suthers and Fattu, 1973; Fattu and Suthers, 1981; Suthers, 1988) with respect to the presented dynamical vocal membrane model are summarized here. Most of the findings were obtained for the bat species *Eptesicus fuscus* that uses frequency modulated (FM) echolocation calls during prey capture and for orientation. Some observations were made on constant frequency (CF) bats.

Vocal membranes in bats are about 6-8 μm thick, 0.5 mm high and 2 mm long. Cuts in vocal membranes left bats aphonic and only able to produce faint clicks. Bat larynges have unusually large cricoid and thyroid cartilages. Their cricothyroid muscle (CTM) is greatly hypertrophied. In bats of the genus *Microchiroptera*, tension of vocal membranes is controlled by the CTM. Electrical activity of the CTM is maximal before pulsed FM vocalizations and then relaxes partially causing the FM downward sweep. CTMs in bats have a contraction time of about 6.5 ms and a total contraction/relaxation time of approx. 12-16 ms. When the motor nerve of the CTM is stimulated at the fusion frequency of 220-240 Hz, the partial relaxation of the CTM vanishes.

In FM bats, the CTM controls both glottal resistance and vocal membrane tension. Glottal resistance corresponds to the aerodynamic pressure drop of the glottis between the vocal folds. The CTM lengthens the vocal folds and vocal

5. Vocal Membrane Model for Bats and Primates

membranes and narrows the glottis. The thyroarytenoid muscle (vocalis muscle) is an additional independent stiffening muscle determining vocal fold visco-elasticity. During pulsed echolocation calls, the pulse type depends on the phase relationship between the contraction/relaxation cycle of the CTM and phonation onset and offset.

Typical pulsed echolocation calls in bats consist of harmonically structured downward sweeps from 60-80 kHz to 30-40 kHz. Pulse duration varies between 1-10 ms, whereas repetition rate can range up to 200 Hz. High sound pressure levels of 120 dB were observed for pulsed FM calls (typical human whisper: 35 dB, human speech: 65 dB, human shout: 90 dB). In experiments with electrically stimulated vocalizations, “protest cries” were observed. Abrupt transitions from calls with discrete FM components to “protest cries” with a wide frequency spectrum were observed. Much of the broadband frequency spectra of such calls lie in the audible range.

Subglottal pressure during pulsed echolocation calls is about 25-40 cmH₂O, peak pressure values reach 60-70 cmH₂O (human conversation 8 cmH₂O, human shouting: 20-30 cmH₂O). Maximal subglottal pressure is limited by the blood pressure in the lung capillaries, about 43 cmH₂O at rest. This is thought to be increased during flight.

Denervation experiments of the intrinsic laryngeal muscles revealed their role in pulsed echolocation calls. The two different branches of the vagus nerve were transected independently: the superior laryngeal nerve which innervates the cricothyroid muscle, and the recurrent laryngeal nerve that innervates all intrinsic muscles except the CTM.

After transection of the recurrent nerve, bats still produce pulsed FM calls. Due to the induced timing problem of the glottal gate, abnormally rising FM pulses were observed. Longer groups of pulses changed to long vocalizations with sinusoidally varying fundamental frequency after transection.

In denervating the cricothyroid muscle, the fundamental frequency of pulsed calls is reduced into the audible range. Thus the frequency modulation of pulsed calls vanishes and pulse series consist of roughly constant frequency pulses at a low fundamental frequency (about 8 kHz). Pulsed calls are still observed; however the pulse duration becomes more variable. In CF bats the effects of CTM transection are similar. Fundamental frequencies of CF calls were lowered from about 83 kHz to approx. 12-42 kHz and more variable from pulse to pulse. Doppler shift compensation disappeared after CTM denervation.

5.1.4 Vocal membranes in squirrel monkeys

In this study, anatomical, neurophysiological and acoustical data from squirrel monkeys (*Saimiri sciureus*), a small New World primate, are used (see Fig. 5.2).

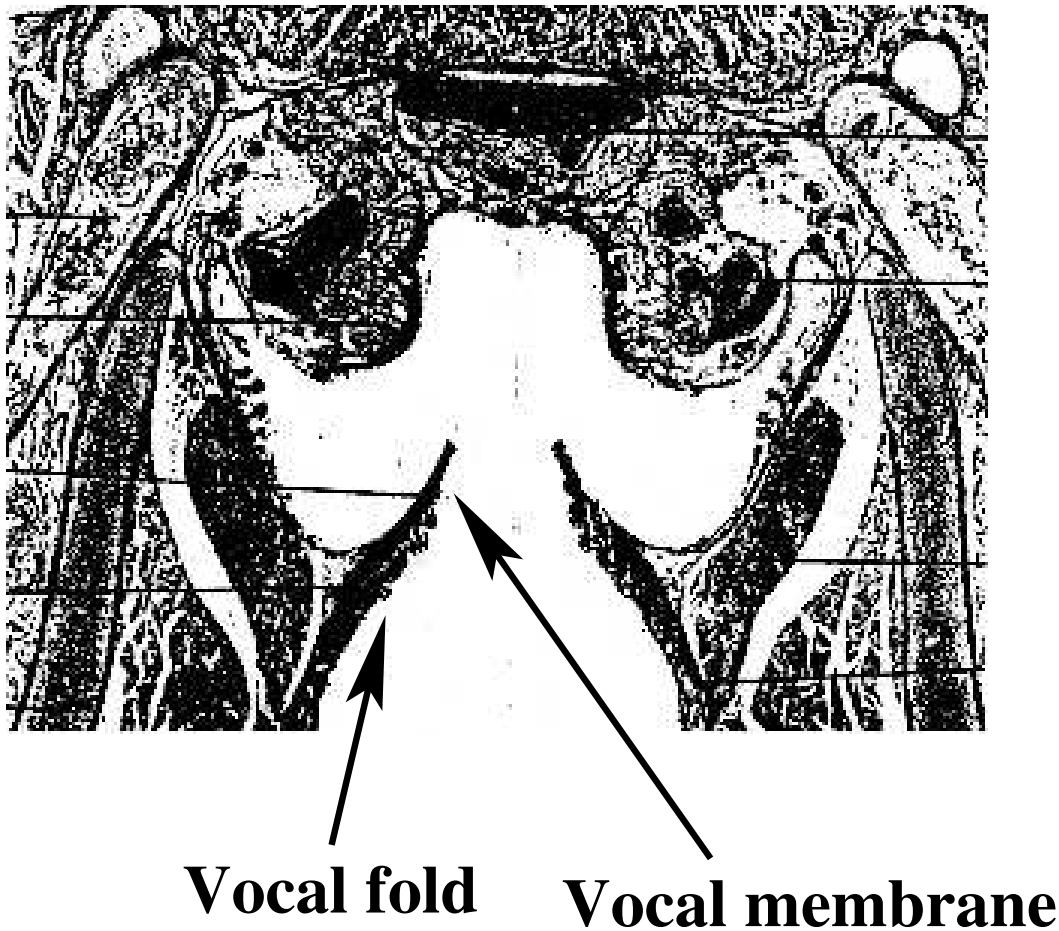


Figure 5.2: Schematic drawing of squirrel monkey larynx (*Saimiri sciureus*) showing vocal folds with vocal membranes (after Starck and Schneider (1960))

Of all primates, vocal behaviors of squirrel monkeys have been studied most intensively (Newman, 1985). Investigations of the sound repertoire and production mechanisms include field studies (Winter, 1972; Newman et al., 1983; Boinski and Mitchell, 1995), laboratory experiments (Winter et al., 1966; Winter, 1969; Schott, 1975; Smith et al., 1982; Häusler, 2000), brain mapping of vocal control centers (Jürgens et al., 1967; Jürgens, 1976a,b, 1988, 1998; Jürgens and Zwirner, 2000), and brain and nerve lesion experiments (Jürgens et al., 1978; Thoms and Jürgens, 1981; Kirzinger and Jürgens, 1985).

The sound repertoire of squirrel monkeys is well documented (e.g., Fig. 5.3), and neural control mechanisms of vocalizations are well known, e.g., in terms of single motor neuron activity. The question on the production mechanism integrating

5. Vocal Membrane Model for Bats and Primates

these experimental observations is still to be answered. Due to the strong conservation of the mammalian phonatory system during evolution, the principles of human voice production may serve as model paradigms for nonhuman mammalian sound production. However, questions on the biomechanics of vocal membranes and their dynamic behavior remain. This study was inspired by a few experimental papers trying to unveil the dynamic significance of vocal membrane in primates. The literature review of these papers can be found in Appendix B.2, page 187.

These studies show that, in general, the vocalization repertoire of squirrel monkeys is highly diverse. On the one hand, it shows a wide range of fundamental frequencies ranging from about 40 Hz up to 16 kHz. On the other hand, there is a large variety of signal complexity from harmonically structured calls to broadband irregular calls (see Fig. 5.3). In particular, the repertoire includes pulsed (purring-like) sounds to high pitched (peep-like) sounds. Squirrel monkey vocalizations contain harmonically structured as well as noisy calls, constant-frequency and frequency-modulated calls, calls with a rather constant amplitude course, calls with marked amplitude modulations, and single syllable calls as well as calls consisting of rhythmic repetitions of specific elements. The important observation is the ability of the aerodynamical-biomechanical system of squirrel monkeys to produce all these patterns within one larynx.

Further details from a few prominent papers (Jürgens et al., 1978; Brown and Cammito, 1995; Brown et al., 2003) relevant for the presented vocal membrane model are summarized here. Transection experiments of intrinsic laryngeal muscles in squirrel monkeys revealed their role in low and high frequency calls. When cutting the recurrent laryngeal nerve, low frequency calls were more effected than high frequency calls. Rhythmical pulse/click series at low pulse rates became arrhythmical. In broadband calls the width of the main spectral energy band decreased or the calls changed to contain harmonic components. Low frequency harmonic calls were superimposed with a broadband spectral component.

Transection of the superior laryngeal nerve (innervating the cricothyroid muscle) left calls with low frequency components unaffected. During shrieking (see Fig. 5.3) faint harmonic-like spectral structures were superimposed on the broadband spectrum. This effect was more prominent during soft shrieks. In calls of a high pitched constant frequency the fundamental frequency decreased and the frequency modulation disappeared. Thus, the CTM is thought to be a major factor in stiffening the vocal membranes. The remaining intrinsic muscles are responsible for pulse production and low frequency calls.

In vivo and *in vitro* studies differentiate vocalization patterns. Two phonatory registers were identified: A modal (chest) register, due to vocal fold and vocal membrane oscillations, is typical at fundamental frequencies of about 600-1400 Hz. A falsetto (loft) register between 1100 and 4800 Hz is generated by vocal

5. Vocal Membrane Model for Bats and Primates

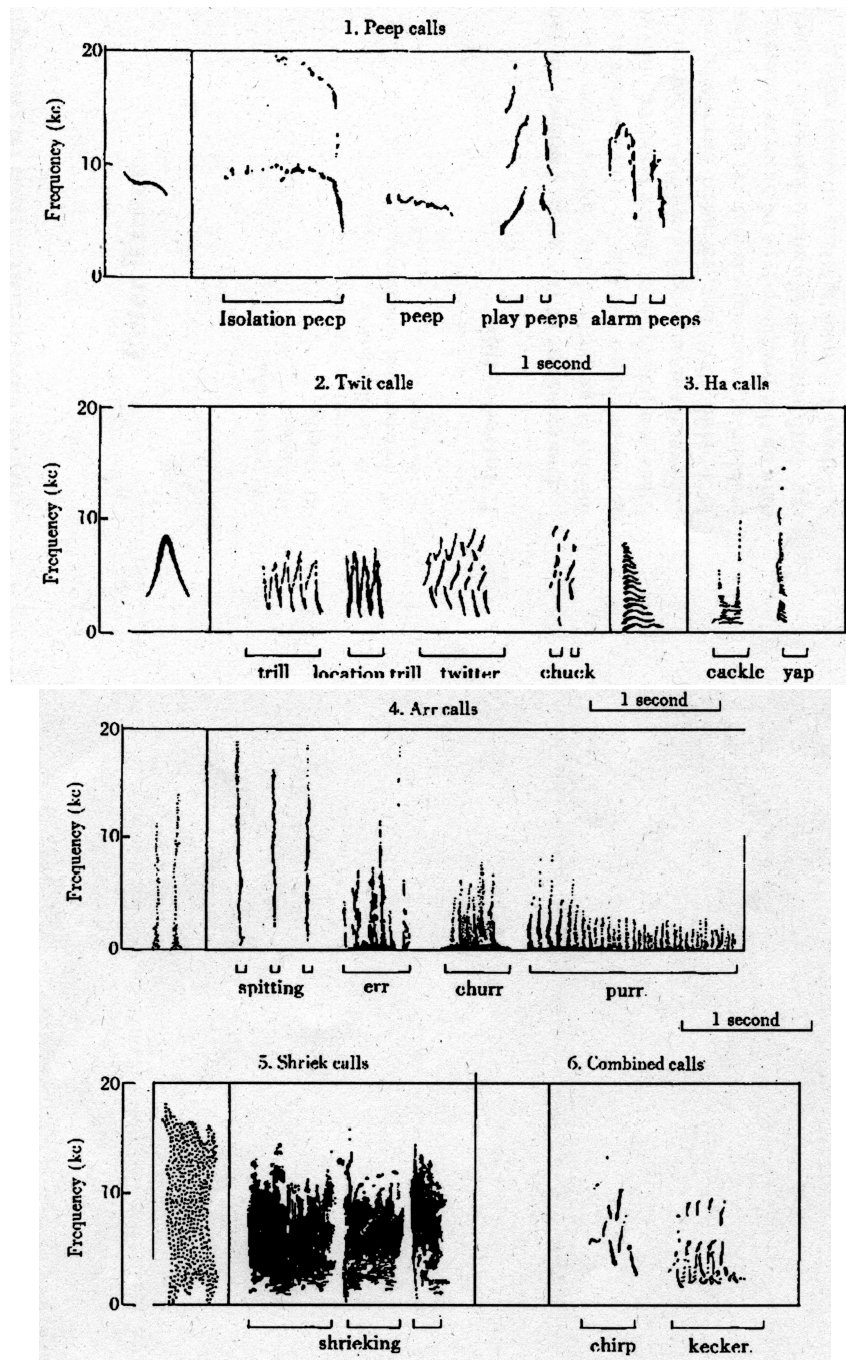


Figure 5.3: Overview of vocalization repertoire of squirrel monkey *Saimiri sciureus*, grouped into “Peep calls”, “Twit calls”, “Ha calls”, “Arr calls”, “Shriek calls”, and “Combined calls” (Winter et al., 1966)

5. Vocal Membrane Model for Bats and Primates

membrane oscillations only. Calls with multiple fundamental frequencies resembling biphonation in human phonation were also found. As is typical for coupled oscillators of this “polyphonic source”, frequency locking and toroidal oscillations with different time courses of different frequency components were observed. In this biphonic regime the vocal membranes and the vocal folds are thought to oscillate in different synchronization modes. *In vitro* experiments with excised larynges showed sustained series of pulsed high frequency oscillations with pulse rates of 8-20 Hz. Furthermore, abrupt transitions between harmonic calls and broadband, irregular phonation, and abrupt jumps between the modal and the falsetto register were observed. The phonatory pressure range could be estimated ranging from approx. 5-8 cmH₂O at onset to about 40 cmH₂O. Hysteresis at onset indicated that phonation is induced by a subcritical Hopf bifurcation from the prephonatory fixed point.

5.1.5 *In vivo* video imaging of oscillating vocal membranes

In recent years, high speed cameras have been used to directly observe vibration patterns of vocal folds in humans (e.g., Farnsworth, 1940; Moore et al., 1962; Kiritani et al., 1993; Berry et al., 2001; Eysholdt et al., 1996). In a previous study, high speed video has been used to observe laryngeal oscillations in squirrel monkeys during brainstem stimulated vocalizations (Fig. 5.4) (Fitch et al., 2001).

Direct evidence could be found for vocal membrane oscillations during high frequency oscillations with fundamental frequencies ranging from 3 to 6 kHz (see Fig. 5.5). Anterior-posterior rocking of the arytenoid cartilages was observed to produce frequency-modulated calls with a modulation frequency of about 15 Hz, such as in twit calls. Details of the temporal behavior of the twit call (e.g., upward or downward pitch movement) depended on the phase relationship between arytenoid rocking and the expiratory cycle. Spectrographs of synchronized acoustical data of brainstem stimulated vocalizations revealed complex temporal and spectral patterns such as abrupt transitions between different oscillatory behaviors and broadband noiselike vibrations (see Figs 5.6).

Due to the low frame rate of the video system, time resolved observations of high frequency oscillations were not possible. Only blurring of high speed video frames could be observed along the glottal midline. This corresponded to blurring of high speed kymograms associated with high frequency phonations (see Fig. 5.5). These recordings imply that tissue oscillations, rather than aeroacoustical whistles, are the primary source for high pitched calls. The dynamical vocal membrane model presented below is based on this conclusion from the high speed video data.

5. Vocal Membrane Model for Bats and Primates

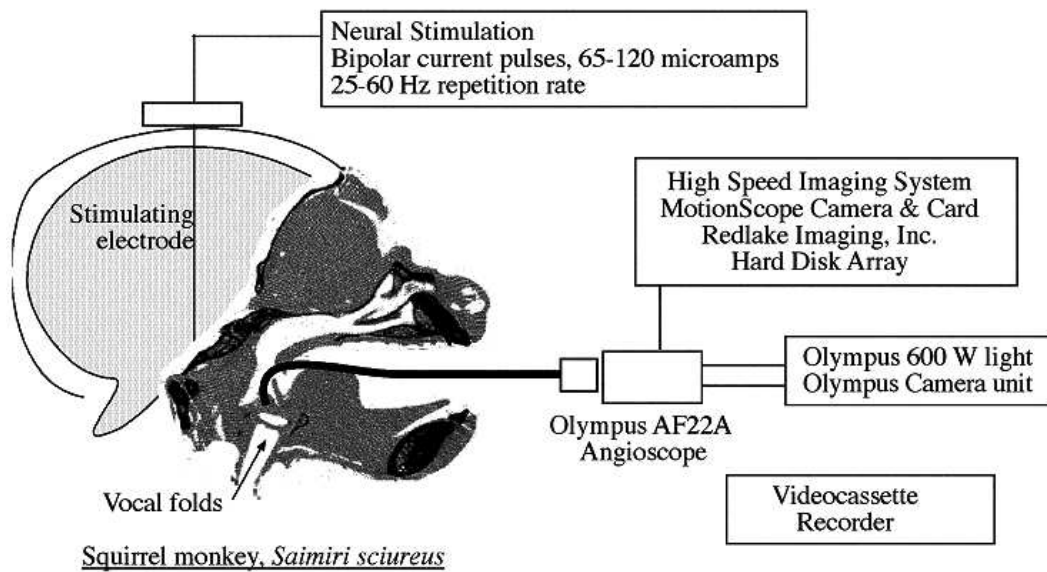


Figure 5.4: Experimental setup for brainstem stimulated squirrel monkey phonation: Vocalizations were elicited by neural stimuli of phonatory motor regions. With a flexible angioscope, the vocal folds were observed and recorded with a digital high speed camera. Frame rates of a high speed imaging system of up to 2000 Hz could be used with the available light system. (Fitch et al., 2001)

5. Vocal Membrane Model for Bats and Primates

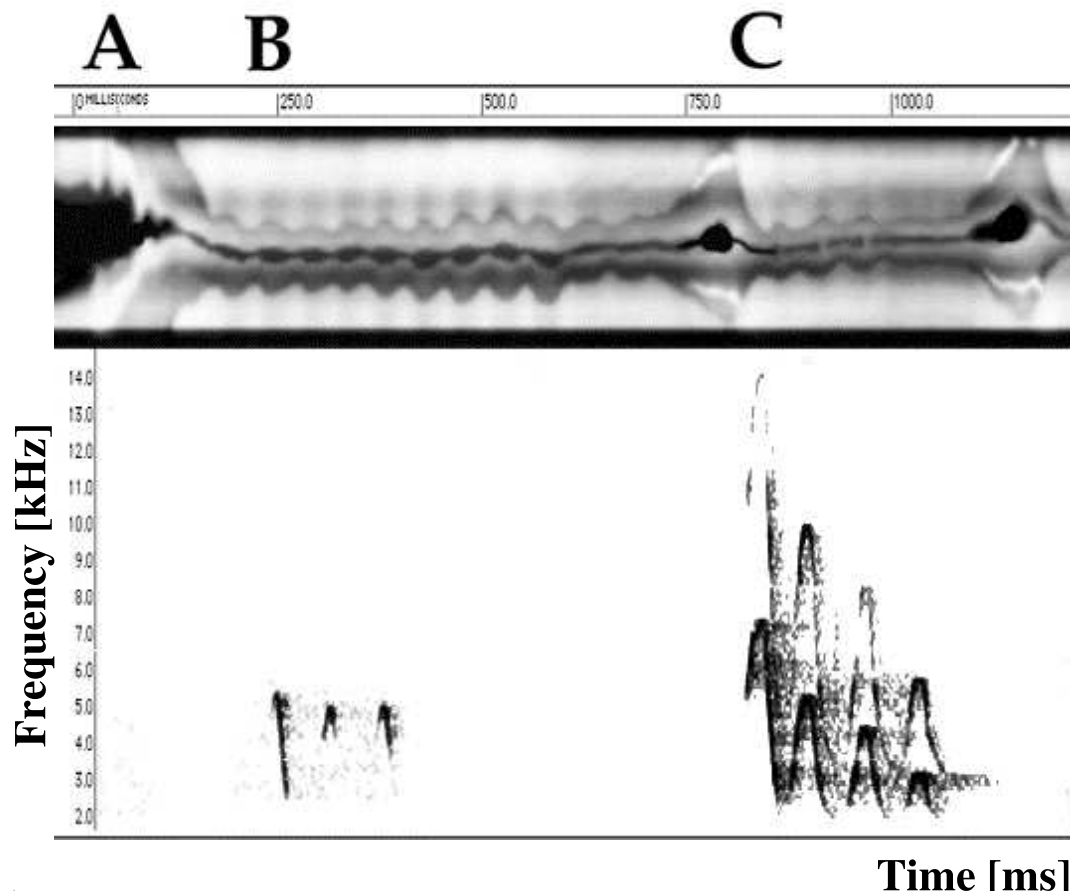


Figure 5.5: Comparison of high speed kymogram (time-space plot for single lateral line across squirrel monkey glottis) with acoustical spectrogram (time-frequency plot) revealing tissue vibrations during high frequency phonation instances and arytenoid rocking modulating the lateral width of the larynx (vocal folds, ventricular folds) thus modulating the fundamental frequency: Segment A: prephonatory closing; Segment B: Arytenoid rocking modulating with of larynx corresponds to frequency modulations in spectrogram; Segment C: After opening of glottis short instances of very high frequency phonation, where the limited temporal resolution of the high speed video camera results in blurred instances in the high speed kymograms. (Fitch et al., 2001)

5. Vocal Membrane Model for Bats and Primates

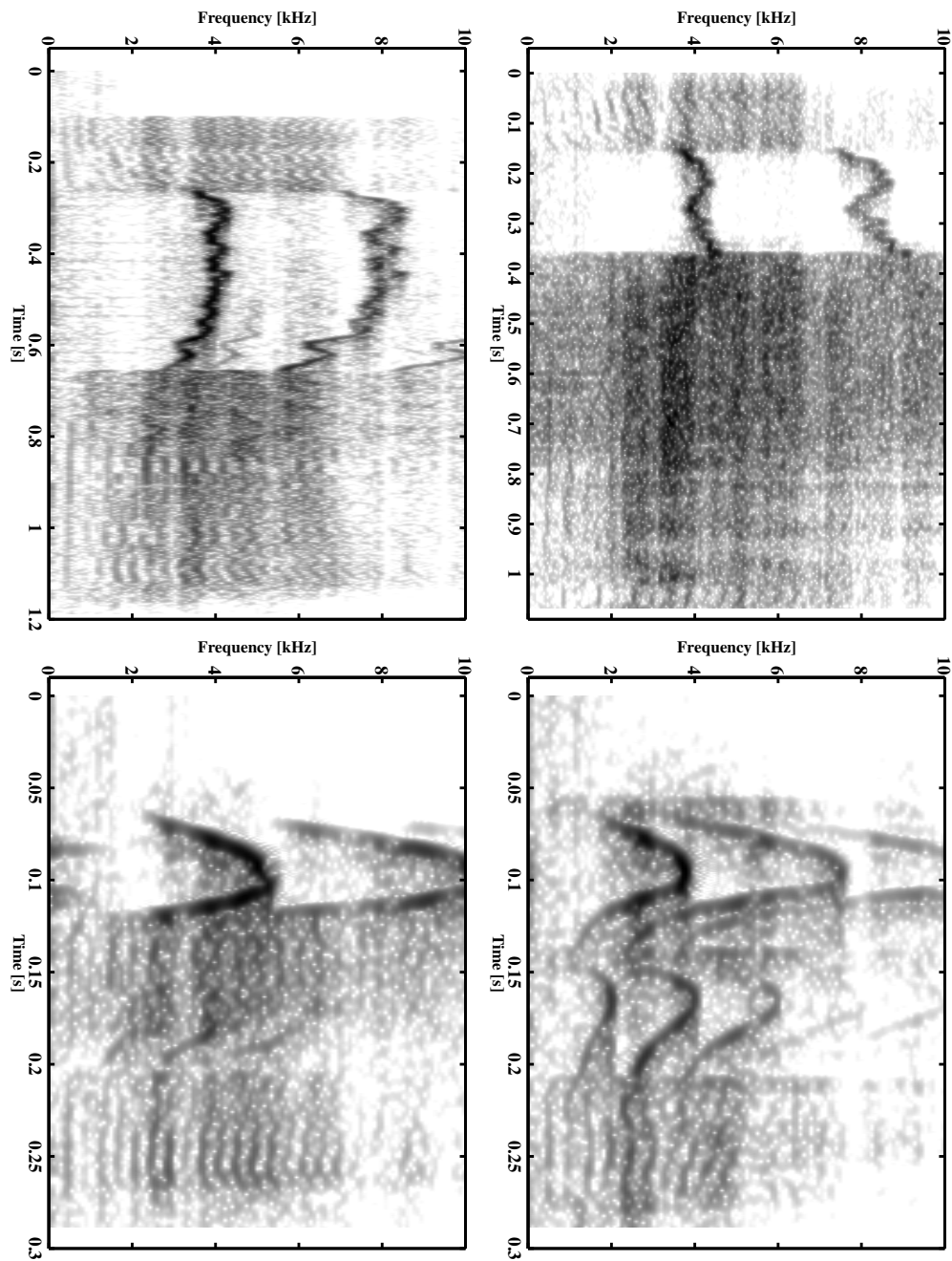


Figure 5.6: Brainstem stimulated vocalizations of a squirrel monkey: rapid transitions from harmonically structured high frequency call segments either to broad-band noiselike behavior or to low frequency harmonic call segments (pitch jump, register jump)

5.1.6 Biomechanical model with rigid reed-like vocal membranes

In a previous study vocal membranes in nonhuman larynges have been modeled by a simplified two-mass model (Steinecke and Herzog, 1995) with upward extensions of rigid massless reeds (Mergell et al., 1999). Using a standard parameter set developed and extensively tested for human larynges, three hypotheses have been addressed:

First, the authors found an optimal membrane geometry where the onset pressure could be minimized (“increased efficiency hypothesis”). Second, for a given subglottal pressure range, the range of oscillation frequencies was significantly enlarged (“increased pitch range hypothesis”). Third, the modified model appeared more susceptible to instabilities that changed the dynamic behavior from simple limit cycle oscillations to a cascade of bifurcations leading to subharmonic oscillations (“chaos hypothesis”). When a large asymmetry between left and right vocal fold (asymmetry factor of upper masses and stiffnesses $q_2 = 0.2$) was introduced in the original two-mass model, simulations with a gradual increase and decrease of subglottal pressure (ranging between 8 and 40 cmH₂O) resulted in only a gradual increase and decrease of fundamental frequency of harmonically structured oscillations. In contrast, simulations of the asymmetrical model with rigid reed-like vocal membranes (having the same asymmetry between left and right vocal fold) showed a cascade of bifurcations to subharmonic oscillations for a gradual increase of subglottal pressure.

5.1.7 Dynamical vocal membrane model

In this chapter a vocal membrane model with dynamic vocal membrane plates will be developed and analyzed. The simplified two-mass model (Steinecke and Herzog, 1995) will be extended by a third mass-spring oscillator attached to the upper edge of the vocal fold. This oscillator resembles the vocal membranes in bats and primates. It is simplified as a rigid plate attached to the superior medial edge of the vocal folds. The connecting hinge incorporates visco-elastic restoring forces. The oscillation frequency of the vocal membrane oscillator is assumed to be orders of magnitudes higher than the typical modal frequencies of the vocal folds.

Observations on the dynamic properties of vocal membranes in bats and primates *in vivo* and *in vitro* will be explained qualitatively by the vocal membrane model. The effects of vocal membranes on *in vacuo* eigenfrequencies and eigenmodes will be studied with linear eigenmode analysis. Phonation onset will be analyzed by Hopf bifurcation analysis. The vocal membrane configuration (membrane mass, height, initial inclination angle) can be optimized to minimize phonation onset pressure and enlarge the phonatory pressure range. In particular, Hopf

bifurcation analysis will reveal the existence of a wide pitch range at phonation onset. It will be shown that a low and a high frequency register exist for the vocal membrane model. Hopf bifurcation analysis will also reveal biphonation in the model. Numerical simulations will show pulsed high frequency oscillations and complex vocalizations with abrupt transitions between different vibratory behavior. Subharmonic oscillations, biphonation, irregular chaotic oscillations and register jumps will be observed in simulations with slowly varying model parameters.

In summary, adding a third oscillator to a simple model for the dynamics of vocal folds, a “nonlinear fit” of vocalization patterns in bats and primates can be achieved. The sound repertoire of bats and primates can be understood qualitatively with a simple low-dimensional model with a small number of parameters. The adaptability of vocal membrane systems in terms of low onset pressure and large phonatory pressure range can be understood even quantitatively.

5.2 Materials and Methods

5.2.1 Phenomenological model of vocal membrane systems

A phenomenological modeling approach was used to study the whole class of non-human larynges with vocal membranes. Although this model lacked a complete microscopic description of anatomical and physiological details of nonhuman larynges, it did reveal basic dynamic behaviors. As in prior studies on nonhuman sound production (on oscine birds (Fletcher, 1988; Fee et al., 1998; Gardner et al., 2001; Laje et al., 2002; Laje and Mindlin, 2002) and on primates and bats (Mergell et al., 1999)), a simple dynamical model was used to mimic vocalization patterns. The phenomenological approach could be regarded as a “nonlinear curve fit”.

The development procedure for such a phenomenological model could be divided up into the following steps:

1. Acoustical data from nonhuman mammals, particularly from squirrel monkeys and bats, were collected and classified to estimate the diversity of nonhuman vocalization repertoires. For signal classification, nonlinear time series analysis was used to identify different dynamic states of the phonatory system and transitions between them. In particular, acoustical data were collected to show the variety of the sound repertoire especially of squirrel monkeys (shown in Fig. 5.3). As shown in Fig. 5.6, the sound repertoire contained rapid transitions between dynamically different states such as limit cycles, subharmonic oscillations, and chaos.
2. Anatomical data from published literature (see, e.g., Fig. 5.1) and morpho-

5. Vocal Membrane Model for Bats and Primates

logical descriptions (e.g., Fig. 5.2) of vocal folds and vocal membranes, in particular of squirrel monkeys, were reviewed and used as a guideline for a schematic qualitative description of the class of vocal membranes in nonhuman mammals (such as in Fig. 5.7).

3. Literature on neural control of phonatory states and transitions between them was reviewed to differentiate between neurally controlled vocal fold behavior and inherent dynamic properties of the myoelastic aerodynamic phonatory system. These data suggested that the phonatory system in nonhuman mammals could be regarded as a biomechanical-aerodynamical system. (Neural control altered the configuration of the nonhuman phonatory system on a time scale that is slower than the dynamical time scale of phonation.) As discussed before, studies on the pulse repetition rate of bats by Griffin et al. (1960) and Suthers and Fattu (1973) failed to uniquely determine the mechanism of high pulse repetition rates of up to 200 Hz. It remains an open question whether the response time of intrinsic muscles to neural stimuli can be fast enough to actively produce a high pulse repetition rate.
4. In the literature the dynamic relevance of vocal membranes was shown in different ways. Indirect methods such as laryngographic waveform measurements and aeroacoustical methods have been used (Brown and Cannito, 1995; Brown et al., 2003). Recently, Fitch et al. (2001) succeeded in directly visualizing oscillating vocal folds and vocal membranes *in vivo* in squirrel monkeys with a high speed video system. Due to limitations of the high speed video system, no time-resolved observations of the oscillating vocal membrane could be made. Nonetheless the observations were sufficient to verify tissue vibrations of the vocal membranes in squirrel monkeys.
5. Following the hypothesis of previous experimental work (Brown and Cannito, 1995; Brown et al., 2003), the model for vocal membranes consisted of two interacting and tunable oscillators. While one oscillator mimicked low frequency oscillations, the other mimics high frequency oscillations. Both oscillators were driven by an airstream provided by lung pressure. A simplified aerodynamic description of the driving airstream coupled left and right side of the bilateral vocal fold and membrane model in a nonlinear way.
6. For model parameter estimation and model evaluation, only robustly measurable data from *in vivo* experiments and observations were used. This included data on fundamental frequencies, physiological driving pressure ranges and observed dynamic patterns.

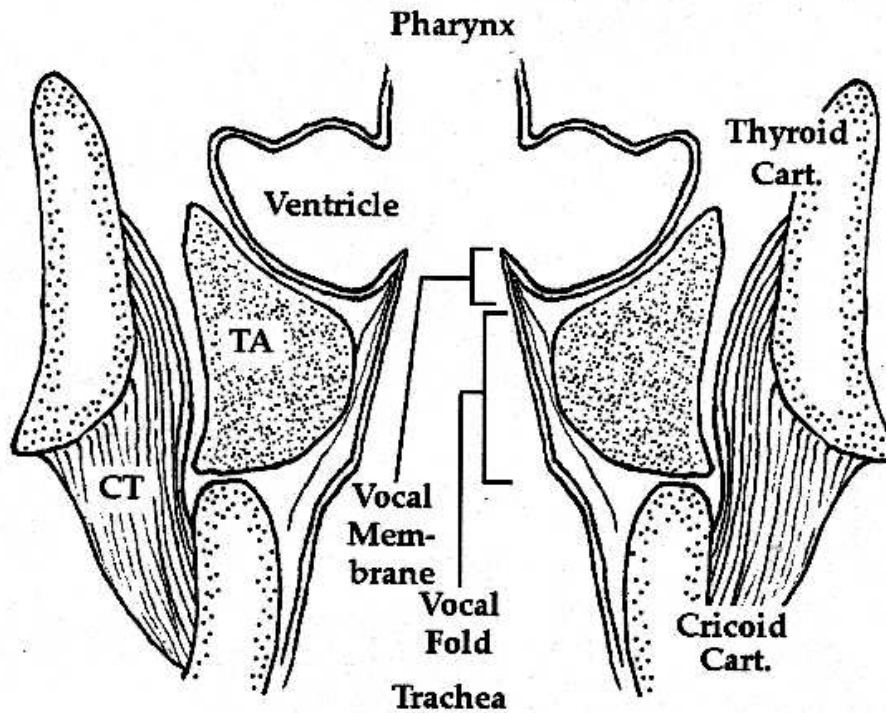


Figure 5.7: Schematic drawing of nonhuman mammalian larynx as viewed through the thyroid cartilage in the front of the larynx. The thyroid cartilage, positioned on the cricoid cartilage, can rotate about a pivot with the cricoid cartilage. Intrinsic muscles, such as the “vocalis muscle”, the thyro-arytenoid (TA), and the cricothyroid muscle (CT) determine the geometrical configuration of the larynx. The vocal folds consist of the vocalis muscle and its membranous cover. The vocal membranes are membranous extensions of the cover tissue into the vocal ventricle. Air from the lungs flowing from the trachea passes the glottis formed by the vocal folds and the vocal membrane. During oscillation, air flow pulses (acoustical waves combined with mean flow) are fed upstream into the trachea and downstream into the ventricle, the pharynx and the upper vocal tract. (from Mergell et al. (1999))

5.2.2 Description of vocal membrane model

Ishizaka and Flanagan (1972) first came up with a biomechanical model for human speech production. This model incorporated the dynamics of the vocal folds and the acoustical filtering of air pulses by the vocal tract. The vocal folds were represented by a damped mass-spring system. The two rectangular-shaped masses on both sides of the bilateral glottis formed a steplike glottal valve. Steinecke and Herzel (1995) modified this original two-mass model to focus on basic dynamic features of the model larynx. Thus vocal tract resonances, viscous losses in the driving airstream and nonlinear restoring forces were neglected.

In a continuum model of the vocal folds, Berry et al. (1994) observed that for normal, regular vocal fold vibrations, two vibrational modes dominated the overall observed variance of excited modes of vibration. These modes were characterized by horizontal (medial) in-phase and out-of-phase oscillations of vocal fold edge points along the vertical medial surface respectively. In the two-mass models, these two modes were mimicked by two box-shaped, horizontally vibrating masses. This neglected modes along the anterior-posterior and inferior-superior direction of the vocal fold surface.

Recent studies showed that the quasi-steady approximation of the flow through the oscillating glottal orifice is valid for most parts of the glottal cycle (Mongeau et al., 1997; Zhang et al., 2002b). The aerodynamic part of the simplified two-mass model was based on this quasi-steady approximation. The driving aerodynamic force was modeled as the aerodynamic pressure produced by the laminar, inviscid and incompressible air flow through the model glottis. This potential flow did not generate self-sustained oscillations of the simplified two-mass model. The further assumption of a jet shedding at the narrowest constriction within the glottis resolved this problem. Jet separation is an inherently viscous and turbulent phenomenon where viscous effects detach the fluid flow from the glottal walls. Details on the shedded vortex structures and on the mixing layer around the jet were neglected in the two-mass model. Fluid flow in the jet downstream of the jet separation point was assumed to be laminar whereas outside of the jet the air is approximated as being at rest.

Here, the studied vocal membrane model was based on the simplified two-mass model. It extended a previous model with massless and rigid reed-like plates as vocal membranes (Mergell et al., 1999). In the vocal membrane model described below, the reed-like plates were assumed to be a third damped mass-spring system that oscillated on top of the simplified two-mass model. For the sake of simplicity, the vocal membrane model was left-right symmetric, and collision forces during contact of left and right vocal membranes were neglected.

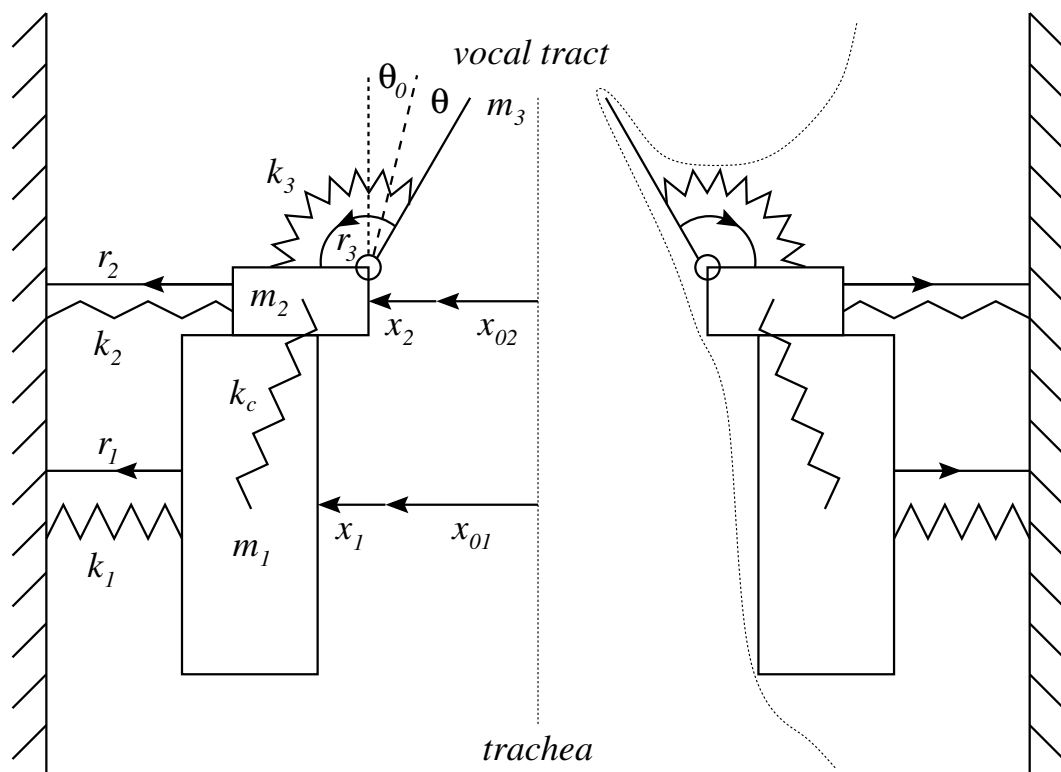


Figure 5.8: Sketch of mechanical model for vocal membrane simulations: Vocal folds are represented by box-like masses (m_1 , m_2) with elastic (k_1 , k_2 , k_c) and viscous (r_1 , r_2) mechanical restoring forces. Vocal membranes are modeled by reed-like plates (with line mass density m_3/d_3) that oscillate about a pivot on top of the upper vocal fold mass. The pivot incorporates elastic (linear elastic force $k_3\theta$ for excursions θ from the resting position θ_0) and viscous (r_3) mechanical restoring forces. The driving air flow, from the trachea upstream of the vocal membrane model to the vocal tract downstream of the model masses and plates, can induce self-sustained oscillations of the model.

5.2.3 Vocal membrane model equations

System of coupled harmonic oscillators

The derivation of the equations of motion follows the sketch of the mechanical model for nonhuman mammalian vocal membrane systems. Details of the derivation of the model equations from Newton's laws can be found in Appendix C.1, page 193.

The dynamics of the model masses was described by a system of coupled har-

5. Vocal Membrane Model for Bats and Primates

monic oscillators. The lower masses (vocal fold) were coupled with each other only by linear elastic restoring forces. Due to the rigid connection of the upper mass and the vocal membrane plate, the upper subsystem contained both dynamic coupling (due to conservation of linear momentum and angular momentum) and aerodynamic coupling (due to forces from the air flow acting on the vocal membrane).

Left and right model masses interacted due to the airstream through the glottal channel and restoring forces during vocal fold contact. The equations of motion are given for one side of the model only as the model is assumed to be left-right symmetric.

The dynamics for the lower mass m_1 of one vocal fold could be written as:

$$m_1\ddot{x}_1 + r_1\dot{x}_1 + k_1x_1 + k_c(x_1 - x_2) = ld_1p_1 + H(-a_1)c_1\frac{|a_1|}{2l} \quad (5.1)$$

with $\dot{x} = \frac{d}{dt}x$ and the Heaviside function $H(a)$. The length of the vocal folds was given by l , the height of the lower masses was d_1 . The cross-sectional area of the glottal channel between the lower masses was $a_1 = 2l(x_1 + x_0)$. The aerodynamic pressure p_1 acted on the walls of the lower glottal channel. During contact of the lower masses, the model masses overlapped. The additional restoring force during contact was proportional to the overlap length $\frac{|a_1|}{2l}$. Thus, after collision, the elastic restoring force $-k_1|x_1|$, ($x_1 < 0$) became $-k_1|x_1| - c_1\frac{|a_1|}{2l}$ which corresponded to an abrupt change of the stiffness constant k_1 .

The equation of motion for the upper masses m_2 – m_3 of one vocal fold were given by:

$$(m_2 + m_3)\ddot{x}_2 + r_2\dot{x}_2 + k_2x_2 + k_c(x_2 - x_1) - \frac{1}{2}m_3d_3\left(\ddot{\theta}\cos(\theta_a) - \dot{\theta}^2\sin(\theta_a)\right) = ld_2p_2 + l \int_0^{d_3\cos(\theta_a)} p(y)dy + H(-a_2)c_2\frac{|a_2|}{2l} \quad (5.2)$$

The height of the upper mass m_2 was d_2 ; the height of the vocal membrane plate was d_3 . θ_a was the absolute vocal membrane angle, as given below. The aerodynamic pressures p_2 and $p(y)$ acted on the walls of the uniform upper glottal channel and on the reed-like vocal membranes respectively. Due to the rigid connection between upper mass m_2 and the vocal membrane plate m_3 , the dynamic coupling term $-\frac{1}{2}m_3d_3\left(\ddot{\theta}\cos(\theta_a) - \dot{\theta}^2\sin(\theta_a)\right)$ appeared. The first part $\ddot{\theta}\cos(\theta_a)$ was due to conservation of linear momentum in the upper subsystem, the second part $\dot{\theta}^2\sin(\theta_a)$ described the centrifugal force of the rotating vocal membrane on the upper mass.

5. Vocal Membrane Model for Bats and Primates

The dynamics for the vocal membranes could be derived as:

$$m_3 \frac{d_3^2}{3} \ddot{\theta} + r_3 \dot{\theta} + k_3 \theta - \frac{1}{2} m_3 d_3 \ddot{x}_2 \cos(\theta_a) = -l \frac{1}{\cos^2(\theta_a)} \int_0^{d_3 \cos(\theta_a)} y p(y) dy \quad (5.3)$$

where the absolute vocal membrane angle was given by:

$$\theta_a(t) = \theta_0 + \theta(t) \quad (5.4)$$

The expression $m_3 \frac{d_3^2}{3}$ was the moment of inertia of the vocal membrane plate with constant line mass density m_3/d_3 with respect to the given pivot point (see Fig. 5.8). The dynamic coupling $-\frac{1}{2} m_3 d_3 \ddot{x}_2 \cos(\theta_a)$ was due to the conservation of angular momentum in the rigidly coupled upper subsystem. The aerodynamic pressure $p(y)$ on the vocal membrane walls resulted in a torque on the vocal membrane.

dynamic coupling of vocal membrane to vocal fold

The coupling terms between the vocal membrane m_3 and the upper mass m_2 of the two-mass system are briefly discussed. Setting the driving pressures to zero, neglecting contact between left and right model masses, and setting the upper masses and the vocal membranes at rest at their resting positions x_{02} and θ_0 ,

$$\dot{x}_2 = x_2 = x_1 = p_i = p(y) = \dot{\theta} = \theta = 0, \quad (5.5)$$

then the equations of motion for the upper subsystem were given by:

1. $(m_2 + m_3) \ddot{x}_2 = \frac{1}{2} m_3 d_3 (\ddot{\theta} \cos(\theta_a) - \dot{\theta}^2 \sin(\theta_a))$: Here the upper mass was accelerated in the positive direction due to an acceleration $\ddot{\theta}$ of the vocal membrane, a consequence of the conservation of linear momentum. For finite angular velocity $\dot{\theta}$, the upper mass was accelerated in the negative direction due to the centrifugal force of the vocal membrane acting on m_2 . Both contributions depended nonlinearly on the absolute angle $\theta_a = \theta_0 + \theta(t)$ of the vocal membrane.
2. $m_3 \frac{d_3^2}{3} \ddot{\theta} = \frac{1}{2} m_3 d_3 \ddot{x}_2 \cos(\theta_a)$: Due to the conservation of angular momentum, the vocal membrane was accelerated in the positive direction by a positive acceleration \ddot{x}_2 of the upper mass.

Aerodynamic driving forces and torques

The vocal membrane model was driven by the aerodynamic pressure from the air flow acting on the walls of the glottal channel. Using Bernoulli's equation, valid along a streamline in the open glottis, the pressures of the quasi-steady, laminar, inviscid and incompressible flow could be derived. A jet was assumed to separate from the glottal walls at the narrowest point in the glottis. Downstream of the jet separation point the air outside of the jet was assumed to be stagnant, whereas the flow inside the jet was assumed to be laminar. Turbulence effects, such as mixing layers and vortex shedding, were neglected. Details on the derivation could be found in Appendix C.2, page 198.

The driving pressure p_1 on the massless plate connected to the lower mass m_1 could be written as:

$$p_1 = p_s \left(1 - \left(\frac{a_{min} H(a_{min})}{a_1} \right)^2 \right) H(a_1) \quad (5.6)$$

The Heaviside functions $H(a)$ assured the open glottis condition for Bernoulli's equation while the glottal area $a_1 = 2l(x_1 + x_{01})$ and the minimum glottal area a_{min} could become negative.

The aerodynamic pressure p_2 on the massless plate connected to the upper mass m_2 read:

$$p_2 = p_s \left(1 - \left(\frac{a_{min} H(a_{min})}{a_2} \right)^2 \right) H(a_1) H(a_2) H(a_1 - a_{vm}) H(a_2 - a_{vm}) \quad (5.7)$$

The Heaviside functions $H(a)$ provided the case differentiations with respect to the glottal areas a_1 , $a_2 = 2l(x_2 + x_{02})$ and the area at the tip of the vocal membranes a_{vm} .

The pressure distribution along the vocal membranes m_3 was given by:

$$p(y) = p_s \left(1 - \left(\frac{a_{min} H(a_{min})}{a_3(y)} \right)^2 \right) H(a_1) H(a_2) H(a_1 - a_{vm}) H(a_2 - a_{vm}) \quad (5.8)$$

The area function $a_3(y)$ for the glottal channel formed by the vocal membranes could be written as:

$$a_3(y) = a_2 - 2l \tan(\theta_a) y, \quad y \in [0, d_3 \cos(\theta_a)] \quad (5.9)$$

The area at the tip of vocal membranes was given by:

$$a_{vm} = a_3(y = d_3 \cos(\theta_a)) = a_2 - 2ld_3 \sin(\theta_a) \quad (5.10)$$

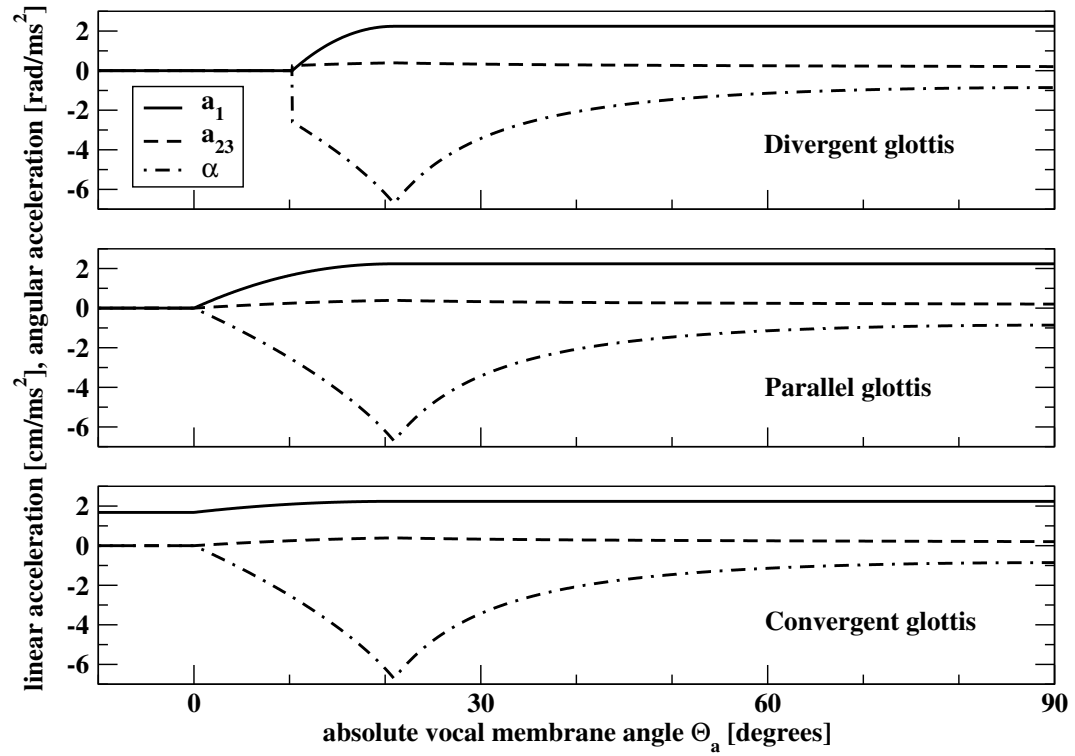


Figure 5.9: Aerodynamic accelerations for different glottal configurations. Vocal membrane parameters were set at $m_3 = 0.25$ and $d_3 = 0.05$, otherwise standard parameter values were used.

5. Vocal Membrane Model for Bats and Primates

Thus, the minimum glottal area could be calculated as:

$$a_{min} = \min(a_1, a_2, a_{vm}) \quad (5.11)$$

Details on the evaluation of the integrals for the aerodynamic force and torque contributions can be found in Appendix C.2, page 198. Here the forces and torque were visualized for different glottal configurations. The forces and torque were scaled by the associated masses and moments of inertia. Therefore, the following linear and angular accelerations were plotted:

$$a_1^{airflow} = \frac{ld_1p_1}{m_1} \quad (5.12)$$

$$a_{23}^{airflow} = \frac{1}{m_2 + m_3} \left[ld_2p_2 + l \int_0^{d_3 \cos(\theta_a)} p(y) dy \right] \quad (5.13)$$

$$\alpha^{airflow} = \frac{3}{m_3 d_3^2} \left[-l \frac{1}{\cos^2(\theta_a)} \int_0^{d_3 \cos(\theta_a)} yp(y) dy \right] \quad (5.14)$$

where $a_1^{airflow}$ denoted the acceleration of the lower mass m_1 , and $a_{23}^{airflow}$ described the acceleration of the upper mass–vocal membrane system m_2 – m_3 . The angular acceleration of the vocal membrane was $\alpha^{airflow}$.

In Fig. 5.9 the aerodynamic accelerations $a_1^{airflow}$, $a_{23}^{airflow}$ and $\alpha^{airflow}$ for different glottal configurations are shown. For these plots, the chosen parameter values were $m_3 = 0.25$, $d_3 = 0.05$. All other parameters were set to their standard values (Table 5.1 on page 122).

1. In the case of a divergent glottis, i.e. divergent vocal folds (e.g., here: $a_1 = 0.025$, $a_2 = 0.05$), all aerodynamic forces vanished until the vocal membrane area $a_{vm}(\theta_a)$ was the minimum area at $\theta_a > 0$. Then, there was a smooth increase of the force on the lower masses. The forces and torques on the upper masses and the vocal membranes jumped abruptly to finite values. When the vocal membrane area was zero, the full subglottal pressure p_s acted on the lower masses. Due to decreasing effective vocal membrane height d_3^{eff} after vocal membrane contact, the forces and (absolute) torques on the upper masses and the vocal membranes decreased gradually.
2. For a parallel glottis (here, e.g., $a_1 = a_2 = 0.05$), all aerodynamic forces smoothly increased for $\theta_a \geq 0$.

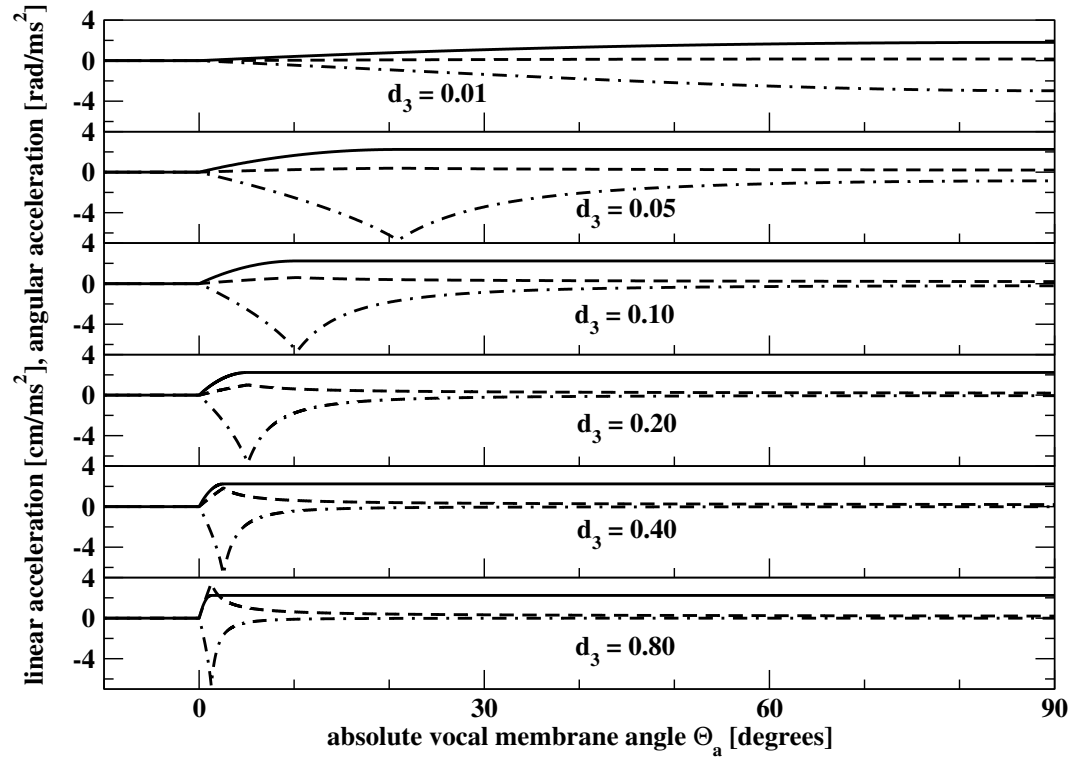


Figure 5.10: Aerodynamic accelerations for different vocal membrane heights. Vocal folds were chosen parallel: $a_1 = a_2 = 0.05$. Vocal membrane parameters were set at $m_3 = 0.25$ and $d_3 = 0.05$, otherwise standard parameter values were used.

5. Vocal Membrane Model for Bats and Primates

3. When the glottis was convergent, i.e. the vocal folds were convergent (e.g., here: $a_1 = 0.10$, $a_2 = 0.05$), the aerodynamic pressure on the lower masses was finite for negative vocal membrane angles $\theta_a < 0$. All other forces and torques on the upper masses and the vocal membranes were zero until the vocal membrane tip area $a_{vm}(\theta_a)$ became the minimum area in the glottis.

Fig. 5.10 shows the influence of the vocal membrane height on the aerodynamic accelerations. The glottis was chosen parallel: $a_1 = a_2 = 0.05$. The vocal membrane mass was $m_3 = 0.025$, all remaining parameters were set to their standard values (see Table 5.1 on page 122). If the vocal membranes were short enough not to be able to close the glottal channel completely (here: $d_3 = 0.01$), the aerodynamic forces and torques changed smoothly with varying vocal membrane angle θ_a . For increasing vocal membrane height the increases of the forces and torques on the upper masses and the vocal membranes for $\theta_a > 0$ became steeper and steeper. The height d_3 had no effect on the driving force for the lower masses m_1 . For very long vocal membranes (here, e.g., $d_3 = 0.80$) the aerodynamic acceleration of the upper mass m_2 could become arbitrarily large, even larger than the acceleration of the lower mass m_1 .

5.2.4 Biologically relevant control parameters

For the analysis of the vocal membrane model a biologically interesting subspace of the parameter space had to be chosen. This choice was guided by scaling the model dynamics and identifying biologically relevant control parameters. The equations of motion for the vocal membrane model were scaled with respect to the accelerated masses.

The dynamics for the lower mass m_1 could be written as:

$$\ddot{x}_1 + \frac{r_1}{m_1} \dot{x}_1 + \frac{k_1}{m_1} x_1 + \frac{k_c}{m_1} (x_1 - x_2) = l d_1 \frac{p_1}{m_1} + H(-a_1) \frac{c_1}{m_1} \frac{|a_1|}{2l} \quad (5.15)$$

Scaling by the sum of the masses of the upper subsystem, $m_2 + m_3$, the equation of motion for the upper masses m_2 – m_3 read:

$$\begin{aligned} \ddot{x}_2 + \frac{r_2}{m_2(1 + \frac{m_3}{m_2})} \dot{x}_2 + \frac{k_2}{m_2(1 + \frac{m_3}{m_2})} x_2 + \frac{k_c}{m_2(1 + \frac{m_3}{m_2})} (x_2 - x_1) - \\ - \frac{1}{2} \frac{d_3}{1 + \frac{m_2}{m_3}} \left(\ddot{\theta} \cos(\theta_a) - \dot{\theta}^2 \sin(\theta_a) \right) = \\ l d_2 \frac{p_2}{m_2(1 + \frac{m_3}{m_2})} + l \frac{1}{m_2(1 + \frac{m_3}{m_2})} \int_0^{d_3 \cos(\theta_a)} p(y) dy + H(-a_2) \frac{c_2}{m_2(1 + \frac{m_3}{m_2})} \frac{|a_2|}{2l} \end{aligned} \quad (5.16)$$

5. Vocal Membrane Model for Bats and Primates

The dynamics of the vocal membrane could be scaled by the moment of inertia of the vocal membrane plate, $m_3 d_3^2/3$:

$$\ddot{\theta} + \frac{3r_3}{m_3 d_3^2} \dot{\theta} + \frac{3k_3}{m_3 d_3^2} \theta - \frac{3}{2} \frac{1}{d_3} \ddot{x}_2 \cos(\theta_a) = -l \frac{3}{m_3 d_3^2} \frac{1}{\cos^2(\theta_a)} \int_0^{d_3 \cos(\theta_a)} yp(y) dy \quad (5.17)$$

Thus, the following biologically relevant control parameters could be introduced:

- Frequency tuning q for the lower two masses: $\omega_i = q \sqrt{\frac{k_i}{m_i}}$, $i = 1, 2$
- Angular frequency of the vocal membranes: $\omega_3^2 = \frac{3k_3}{m_3 d_3^2}$
- Mass ratio of the upper mass m_2 and the vocal membrane mass m_3 : $q_m = \frac{m_2}{m_3}$
- Geometry parameters: vocal membrane height d_3 , resting angle θ_0

All remaining parameters were adopted from the simplified two-mass model (Steinecke and Herzel, 1995) for comparability. The parameters for the simplified model have been chosen to fit the dynamics to observed time series obtained from human phonation.

It should be noted that for fixed mass ratio q_m and for scaled lower and upper masses $m_i^{scaled}(q) = m_i/q$, $i = 1, 2$ the vocal membrane mass depended on the frequency tuning: $m_3 = m_2/(q_m q)$. If the angular frequency ω_3 of the vocal membranes was fixed, the stiffness constant of the vocal membrane was scaled with q : $k_3 = \frac{1}{3} m_3 d_3^2 \omega_3^2$.

Mass ratio limits $q_m \rightarrow 0$ and $q_m \rightarrow \infty$

Two limit cases for the mass ratio q_m are discussed in the following. It will be shown that q_m could be interpreted as a coupling parameter between the two-mass subsystem m_1 – m_2 (representing the vocal folds) and the vocal membrane m_3 .

For $q_m \rightarrow \infty$, i.e. $m_2 \gg m_3 > 0$ the two-mass subsystem was coupled to the vocal membranes only by Bernoulli forces. This forces acted on the nearly massless reed-like vocal membranes on top of the upper mass m_2 . In the equation of motion for m_2 the influence of the dynamic coupling term $\ddot{\theta} \cos(\theta_a) - \dot{\theta}^2 \sin(\theta_a)$ vanished as $q_m \rightarrow \infty$; whereas the denominator approached $m_2(1 + \frac{1}{q_m}) \rightarrow m_2$. In this case the previous model with rigid massless reed-like plates by Mergell et al. (1999) was recovered.

5. Vocal Membrane Model for Bats and Primates

For $q_m \rightarrow 0$, i.e. $0 < m_2 \ll m_3$, the equation for the upper mass m_2 could be written as:

$$\frac{1}{m_3(1+q_m)} = \frac{q_m}{m_2(1+q_m)} \rightarrow 0 \text{ for } q_m \rightarrow 0 \quad (5.18)$$

$$\Rightarrow \ddot{x}_2 = \frac{1}{2}d_3 \left(\ddot{\theta} \cos(\theta_a) - \dot{\theta}^2 \sin(\theta_a) \right) \quad (5.19)$$

In this case the dynamics of the upper mass m_2 was entirely determined by the dynamics of the vocal membrane m_3 . The vocal membrane interacted with the upper mass due to $\ddot{\theta} \cos(\theta_a)$ (the expression for conservation of momentum) and the centrifugal force $-\dot{\theta}^2 \sin(\theta_a)$. The dynamics of the vocal membrane became:

$$\begin{aligned} \left(1 - \frac{3}{4} \cos^2(\theta_a)\right) \ddot{\theta} + \left(\frac{3r_3}{m_3 d_3^2} + \frac{3}{4} \sin(\theta_a) \cos(\theta_a) \dot{\theta}\right) \dot{\theta} + \frac{3k_3}{m_3 d_3^2} \theta = \\ -l \frac{3}{m_3 d_3^2 \cos^2(\theta_a)} \int_0^{d_3 \cos(\theta_a)} yp(y) dy \end{aligned} \quad (5.20)$$

Thus, for negative angular velocities, $\dot{\theta}(t) < 0$, during opening of the vocal membranes, the effective damping r_3^{eff} of the vocal membrane was reduced.

$$r_3^{eff} = \left(\frac{3r_3}{m_3 d_3^2} + \frac{3}{4} \sin(\theta_a) \cos(\theta_a) \dot{\theta} \right) \quad (5.21)$$

Note that r_3^{eff} could even become negative. If the damping was negative, the subsystem upper mass–vocal membrane could exhibit limit cycle oscillations. Furthermore, when the vocal membranes moved outside the driving airstream ($\theta_a(t) = \theta_0 + \theta(t) < 0$), the effective damping was reduced to possibly negative values.

5.2.5 Numerical analysis of the vocal membrane model

The set of differential equations of second order for the coupled harmonic oscillators of the vocal membrane model could be transformed to a set of first-order differential equations. A set of new variables was introduced to rewrite the vocal membrane equations as a set of inhomogenous first order differential equations with nonlinear coupling:

$$v_1 := \dot{x}_1, v_2 := \dot{x}_2, \omega := \dot{\theta} \quad (5.22)$$

This set of first-order equations was integrated numerically. Bifurcation analysis was performed with the software tool XPP/XPPAUT which includes the bifurcation analysis tool AUTO.

The set of standard parameters was adopted from the simplified two-mass model developed for normal human male phonation (Steinecke and Herzel, 1995).

5. Vocal Membrane Model for Bats and Primates

m_1	m_2	q_m	k_1	k_2	f_3	k_c	r_1	r_2	R_3	l
0.125	0.025	10.0	0.08	0.008	1.0	0.025	0.02	0.02	1.0	1.4
c_1	c_2	a_{01}	a_{02}	θ_0	d_1	d_2	d_3	ρ	p_s	
$3k_1$	$3k_2$	0.05	0.05	1.0	0.25	0.05	0.05	0.00113	0.008	

Table 5.1: Set of standard parameters for the vocal membrane model adopted from the set of standard parameters for the simplified two-mass model for normal male human phonation (in the unit system cm, ms, g, degree)

Additionally, standard parameters for the vocal membrane plate were added. The vocal membrane height d_3 and the resting angle θ_0 were chosen to be comparable with values used by Mergell et al. (1999). There $d_3 = 0.1$ cm and $\theta_0 = 1.8$ degrees were found to be optimal for minimum onset pressure. Using the biologically relevant control parameters described above, the complete set of parameters is given in Table 5.1.

In Table 5.1, the resting areas for the masses m_1 and m_2 were defined as:

$$a_{01} = a_{01l} + a_{01r} \quad (5.23)$$

$$a_{02} = a_{02l} + a_{02r} \quad (5.24)$$

The mass m_3 of the vocal membrane was given in terms of the mass ratio q_m :

$$m_3 = \frac{m_2}{q_m} \quad (5.25)$$

The damping constant r_3 was given by:

$$r_3 = R_3 \frac{1}{3} m_3 d_3^2 \quad (5.26)$$

The damping constant could be compared to the critical damping known from linear damped harmonic oscillators. Here the critical damping of the vocal membrane without driving forces and without coupling to the upper mass m_2 was given by:

$$r_3^{crit} = 2\sqrt{\frac{1}{3}k_3 m_3 d_3^2} \quad (5.27)$$

Combined with the angular frequency of the vocal membrane plate,

$$\omega_3^2 = \frac{3k_3}{m_3 d_3^2} = (2\pi f_3)^2 \quad , \quad (5.28)$$

5. Vocal Membrane Model for Bats and Primates

(where the stiffness constant k_3 is related to the frequency f_3), then the damping ratio $\zeta_3 = r_3/r_3^{crit}$ could be written as:

$$\zeta_3 = \frac{R_3}{2\omega_3} \quad (5.29)$$

For the standard parameters of the vocal membrane model, the damping ratio $\zeta_3 \approx 0.08 \leq 1$ indicated subcritical damping of the vocal membrane plate. This was consistent with damping ratios found in rheometrical stress-strain measurements of human vocal fold tissues (Chan and Titze, 1999, 2000; Chan, 2001) and obtained from *in vivo* resonance experiments on human vocal folds (Švec et al., 2000).

It is known that the eigenfrequencies of the simplified two-mass model are around 100 Hz (Berry, 2001). The frequency $f_3 = 1000 \text{ Hz} = \frac{1}{ms}$ was chosen one order of magnitude higher than the typical oscillation frequency of the two-mass model. Aiming at modeling high frequency phonation in bats and nonhuman primates, this frequency was sufficiently higher than the eigenfrequencies of the vocal fold part of the model. This allowed the study of the interaction of the tunable low frequency vocal fold oscillations and the high frequency vibrations of the vocal membrane. Furthermore, this frequency choice was appropriate for efficient and fast numerical integration of the vocal membrane model.

The model equations were integrated with a numerical 4th order Runge-Kutta scheme. The step size was fixed to a sampling rate of $f_s = 800 \text{ kHz} = 800 \frac{1}{ms}$. An integrator scheme with adaptive step size failed due to discontinuities of the derivative of the driving forces. The step size had been varied to find good convergence of the numerical integration scheme. Numerical tests with integrator schemes in XPP/XPPAUT for stiff differential equations confirmed the used step size.

The standard initial conditions for the numerical integration were:

$$\begin{aligned} x_{1l}(t_0) &= x_{1r}(t_0) = 0.01 \\ x_{2l}(t_0) &= x_{2r}(t_0) = 0.0 \\ \theta_l(t_0) &= \theta_r(t_0) = 0.0 \\ v_{1l}(t_0) &= v_{1r}(t_0) = 0.01 \\ v_{2l}(t_0) &= v_{2r}(t_0) = 0.0 \\ \omega_l(t_0) &= \omega_r(t_0) = 0.0 \end{aligned} \quad (5.30)$$

$$\quad (5.31)$$

with the subscripts “l” and “r” indicating left and right vocal fold and membrane mass displacements.

For the numerical evaluation of the Heaviside function $H(a)$, the approximation

$$H(a) := \begin{cases} 0 & , a < 0 \\ \tanh(10^6 a) & , a \geq 0 \end{cases} \quad (5.32)$$

was used.

5.3 Results

5.3.1 Linear eigenmode analysis

In this section, the results of linear eigenmode analysis of the vocal membrane model are described and briefly discussed. The influence of the introduced control parameters on eigenfrequencies and amplitudes and phase angles of the eigenmodes was studied.

The ranges of the tuning parameter q , where entrained oscillations of the full model could be expected, depended on the mass ratio q_m , the resting angle θ_0 , and the vocal membrane height d_3 . Comparable eigenfrequencies were expected to be related to frequency locking, toroidal oscillations and chaos in the full nonlinear model. Eigenmode analysis allowed to classify oscillatory modes as “vocal fold modes” and “vocal membrane modes”. These different modes could facilitate the definition of different registers of the vocal membrane model. The study of eigenmode amplitudes and phase relations, depending on the geometrical and dynamical control parameters d_3 , θ_0 , and q_m , could help to qualitatively understand the phonation onset behavior of the full nonlinear vocal membrane model.

For the study of the eigenfrequencies and eigenmodes of the vocal membrane model, the linearized homogeneous equations of motion without driving forces could be written as:

$$m_1\ddot{x}_1 + r_1\dot{x}_1 + k_1x_1 + k_c(x_1 - x_2) = 0 \quad (5.33)$$

$$\begin{aligned} \left(1 - \frac{3 \cos^2(\theta_0)}{4(1 + q_m)}\right) \left(1 + \frac{1}{q_m}\right) m_2\ddot{x}_2 + r_2\dot{x}_2 + (k_2 + k_c)x_2 - k_cx_1 + \\ + \frac{3 \cos \theta_0}{2} \frac{1}{d_3} [r_3\dot{\theta} + k_3\theta] = 0 \end{aligned} \quad (5.34)$$

$$\begin{aligned} \left(1 - \frac{3 \cos^2(\theta_0)}{4(1 + q_m)}\right) \frac{1}{3} m_3 d_3^2 \ddot{\theta} + r_3\dot{\theta} + k_3\theta + \\ + \frac{1}{2} \frac{d_3 \cos(\theta_0)}{1 + q_m} [r_2\dot{x}_2 + (k_2 + k_c)x_2 - k_cx_1] = 0 \end{aligned} \quad (5.35)$$

A complex exponential was chosen as an ansatz for the solution of the linearized homogeneous vocal membrane equations:

$$\begin{pmatrix} x_1 \\ x_2 \\ \theta \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \\ A_\theta \end{pmatrix} e^{\lambda t} \quad (5.36)$$

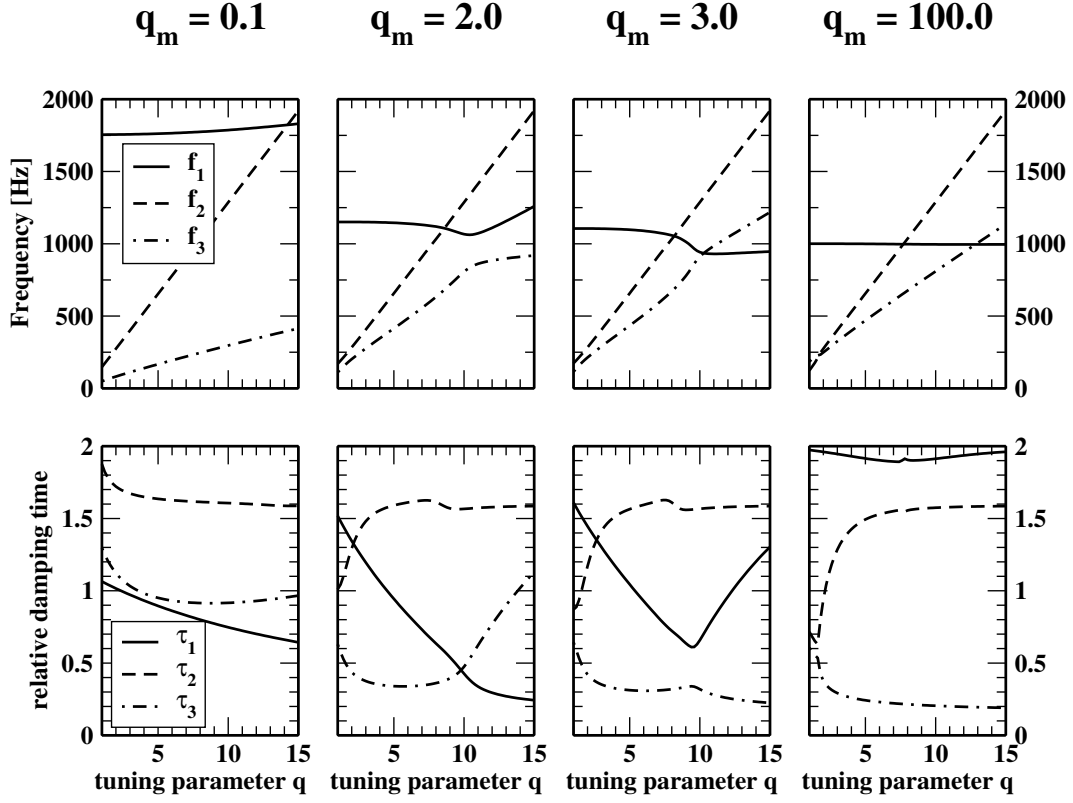


Figure 5.11: Variation of eigenfrequencies and relative damping times with mass ratio q_m where $d_3 = 0.05$ cm and $\theta_0 = 0$ degree, otherwise standard parameter values were used.

where $A_1, A_2, A_\theta \in \mathbb{C}$ were complex amplitudes, and $\lambda \in \mathbb{C}$ was the complex frequency .

The solvability condition for the algebraic set of linear equations in the complex amplitudes lead to the characteristic polynomial, a 6th order polynomial for the complex frequency λ . The complex solutions $\lambda_i = \gamma_i \pm i\omega_i$, $i = 1, 2, 3$ lead to the eigenfrequencies $f_i = \frac{\omega_i}{2\pi}$ and the damping time $\tau_i = 1/\gamma_i$. The ratio of the damping time τ_i to the eigenperiod $1/f_i$, termed the relative damping time $\tau_i^{rel} = f_i/\gamma_i$, was used in the following analysis. The moduli $|A_i|$ and the differences of the phase angles φ_i , $i = 1, 2, \theta$ of the complex amplitudes A_i , $i = 1, 2, \theta$ were calculated with the solutions λ_i of the characteristic polynomial.

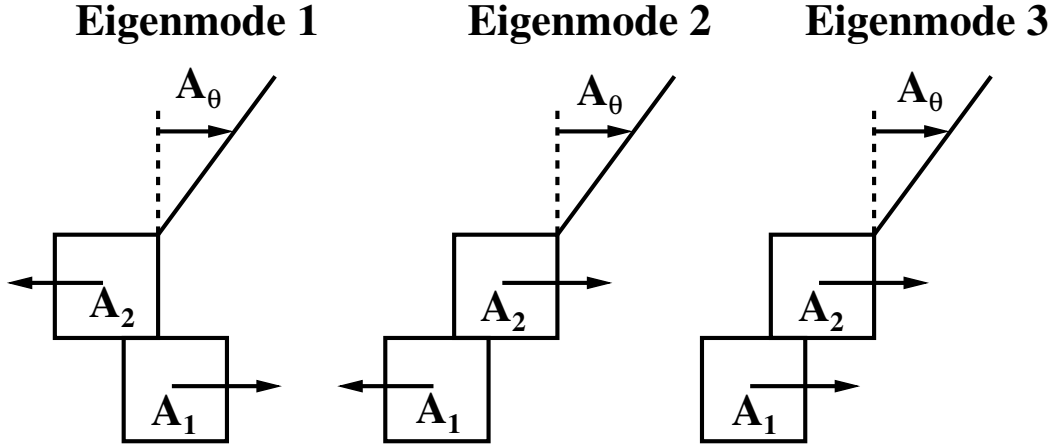


Figure 5.12: Typical shape of eigenmodes of the vocal membrane model

Influence of mass ratio q_m

In Fig. 5.11 for four different mass ratios $q_m \in \{0.1, 2, 3, 100\}$, the eigenfrequencies and the relative damping times as functions of the tuning parameter q are shown. It was observed that for small mass ratios (e.g., $q_m = 0.1$) the first eigenfrequency f_1 shifted towards higher frequencies. The relative damping time τ_1 is mostly smaller than one.

This eigenmode had a large vocal membrane component whereas the amplitudes of the two lower masses were very small. The phase relations of the first eigenmode are shown in Fig. 5.12, where all masses of the model oscillated typically out-of-phase. This mode is termed “vocal membrane mode”. The second eigenfrequency was characterized by large eigenmode amplitudes of the lower mass m_1 and the vocal membrane. Fig. 5.12 shows typical phase relations of the second mode. Out-of-phase oscillation of the lower and upper mass and in-phase oscillation of the upper mass and the vocal membrane mass characterized the second mode (Fig. 5.14). The third eigenfrequency corresponded to large eigenmode amplitudes of the upper mass m_2 and the vocal membrane. As shown in Fig. 5.12, the third eigenmode consisted of in-phase oscillations of all masses of the vocal membrane model (Fig. 5.14). The second and third mode were called “vocal fold modes”. They could be subdivided further into “glottal wave mode” for the second eigenmode and “vocal fold bulk mode” for the third eigenmode.

For small mass ratios $q_m = 0.1$, the variation of the first and the third eigenfrequencies, f_1 , and f_3 , with increasing tuning parameter q were small in comparison to the variation of the second eigenfrequency f_2 . Frequency locking at comparable frequencies was expected only at high values of q , approx. $q \geq 13$.

5. Vocal Membrane Model for Bats and Primates

As the mass ratio increased, the distance between the first eigenfrequency f_1 and the third eigenfrequency f_3 decreased, until $f_1(q)$ and $f_3(q)$ intersected. For increasing mass ratio, the damping time τ_1 increased; whereas τ_3 decreased. The region for entrained oscillations of the full nonlinear model could be expected to be in the interval $6 \leq q \leq 10$.

For large mass ratios (e.g., $q_m = 100$), $f_1(q)$ was nearly constant and approached the value corresponding to the fixed angular frequency $\omega_3 = 2\pi 1000\text{Hz}$. The relative damping time τ_1 became the largest damping time. The interaction with eigenmodes 2 and 3 vanished as the third eigenfrequency became a linear function of the tuning parameter. The relative damping time τ_3 became the smallest damping time. From the theory of coupled oscillators it could be expected that mode entrainment could occur in the intervals $6 \leq q \leq 10$ and $12 \leq q \leq 15$.

Varying the mass ratio q_m , the second eigenfrequency f_2 showed no significant changes. The effect on the damping time τ_2 was also very small. Furthermore, no significant effect of q_m on the eigenmodes could be observed.

Within the framework of coupled oscillators (Bergé et al., 1984), the mass ratio q_m can be expected to be a crucial control parameter influencing the frequency ratio of the coupled oscillators. It has a significant influence on the distance between eigenfrequencies as a function of the tuning parameter q . The mass ratio q_m varies the ranges of the tuning parameter q where entrained oscillations of the full nonlinear model can be expected.

Furthermore, an increasing mass ratio was shown to increase the relative damping time of the vocal membrane mode. This could affect the onset behavior of self-sustained vocal membrane oscillations as viscous energy dissipation is reduced. It could also be crucial for only weakly damped high frequency oscillations during low frequency oscillation of the vocal fold masses. This effect could be crucial for the generation of pulsed echolocation calls in bats (Fattu and Suthers, 1981; Suthers and Fattu, 1973).

Influence of resting angle θ_0

In Fig. 5.13, the variation of the eigenfrequencies and the relative damping time with increasing resting angle θ_0 is shown. The chosen mass ratio $q_m = 2.0$ allowed comparison with the eigenfrequencies and relative damping times for $\theta_0 = 0$ degree, shown in Fig. 5.11.

As the resting angle increased, a decrease of the distance between the first and the third eigenfrequency could be observed. At $\theta_0 = 30$ degrees, $f_1(q)$ and $f_3(q)$ intersected. For large θ_0 (e.g., $\theta_0 = 60$ degrees), $f_1(q)$ and $f_3(q)$ were nearly linear functions of the tuning parameter q .

The relative damping time τ_1 of the vocal membrane mode was shifted to higher values for increasing resting angles θ_0 . The relative damping time τ_3 was lowered;

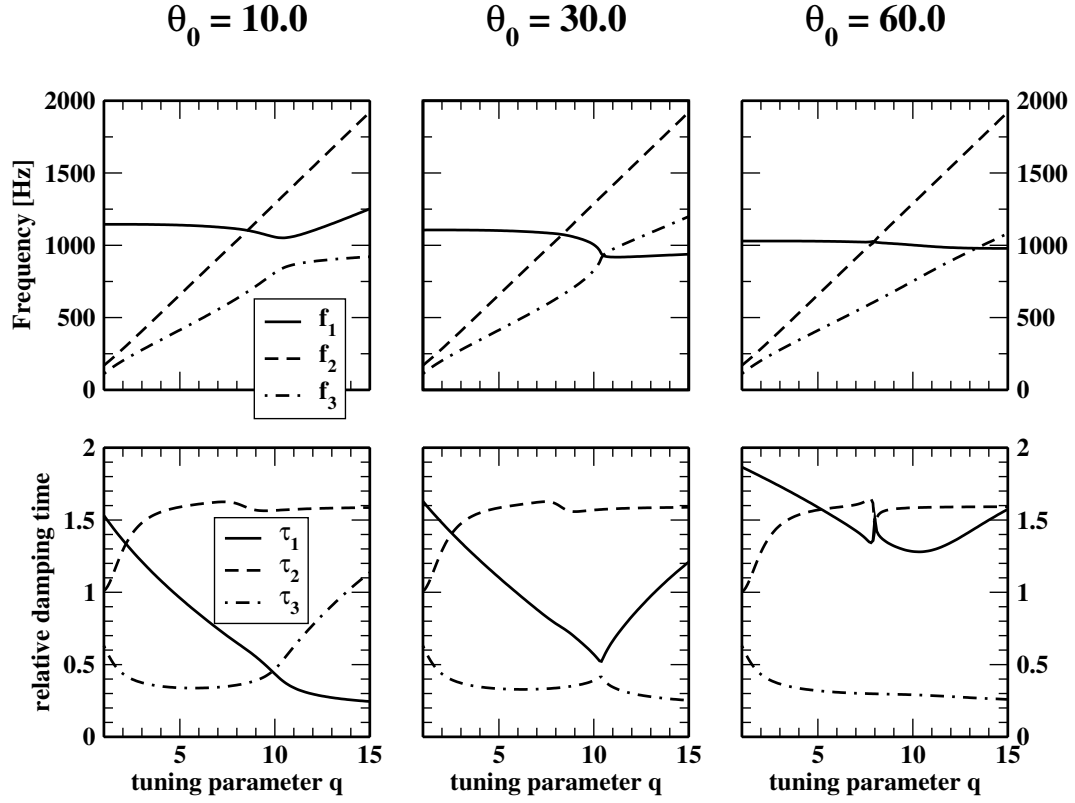


Figure 5.13: Variation of eigenfrequencies and relative damping times with resting angle θ_0 [degrees] where $d_3 = 0.05$ cm and $q_m = 2.0$, otherwise standard parameter values were used.

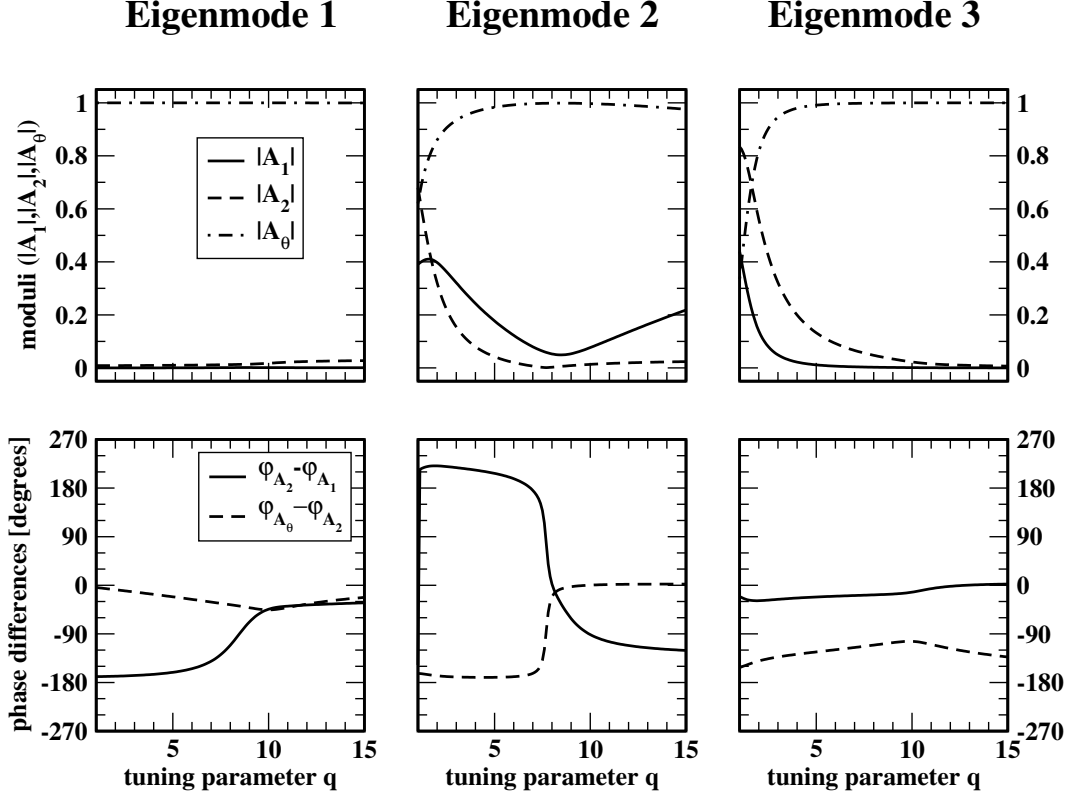


Figure 5.14: Eigenvectors of the vocal membrane model for $d_3 = 0.05$ cm, $q_m = 2.0$, $\theta_0 = 0$ degree, otherwise standard parameter values. Note that due to the definition of the vocal membrane angle θ , a zero phase difference $\varphi_{A_\theta} - \varphi_{A_2} = 0$ between the upper mass and the vocal membrane corresponds to, e.g., an increase of the glottal area a_2 and a decrease of the vocal membrane area a_{vm} .

whereas the relative damping time τ_2 remained unaffected by varying θ_0 .

No significant effect of varying resting angles θ_0 on the eigenmodes could be observed. They remained similar to the eigenmodes shown in Fig. 5.14.

Similar to the influence of the mass ratio q_m , the resting angle θ_0 could be expected to influence the ranges of the tuning parameter q , where entrained oscillations of the full nonlinear vocal membrane model could occur. Increasing resting angles also decreased the damping of the vocal membrane mode. For self-sustained vocal membrane oscillations and pulsed oscillations (e.g., pulsed echolocation calls in bats) the resting angle could facilitate these modes of vibration.

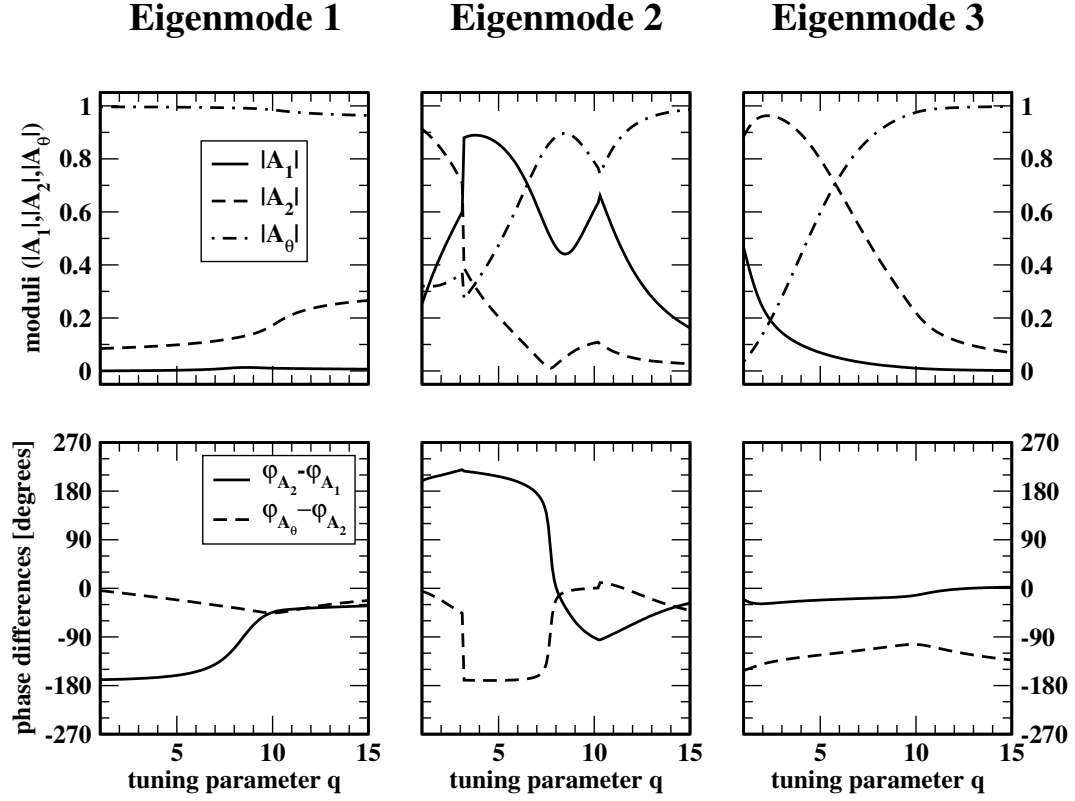


Figure 5.15: Eigenvectors of the vocal membrane model for $d_3 = 0.50$ cm, $q_m = 2.0$, $\theta_0 = 0$ degree. Otherwise standard parameter values were used.

Influence of vocal membrane height d_3

In Figs 5.14 and 5.15, the eigenmodes for two different vocal membrane heights, $d_3 = 0.05$ cm and $d_3 = 0.5$ cm, are shown. The eigenfrequencies and relative damping times for these cases were similar to the results shown in Fig. 5.11.

For small vocal membrane height d_3 (e.g., $d_3 = 0.05$ cm for Fig. 5.14), the first eigenmode (“vocal membrane mode”) was characterized by a large vocal membrane component. The components of the lower and upper masses were about two orders of magnitudes smaller. As a function of the tuning parameter q , the two lower masses changed from oscillating out-of-phase for small q to vibrating in-phase at larger q . The upper mass m_2 and the vocal membrane m_3 showed in-phase oscillations for the shown range of tuning parameter variations. Note that due to the definition of the vocal membrane angle θ (Fig. 5.8), in-phase oscillation of the upper mass and the vocal membrane corresponded to out-of-phase oscillation of the glottal area a_2 and the area at the tip of the vocal membrane a_{vm} .

For larger vocal membrane heights d_3 (e.g., $d_3 = 0.5$ cm for Fig. 5.15) the eigenmode component of the upper mass was significantly increased. The eigenmode component of the lower mass remained small as for small vocal membrane heights. No significant effect on the phase differences could be observed.

The second eigenmode changed more drastically with increasing vocal membrane height. For small vocal membrane height, the components of the lower mass m_1 and the vocal membrane m_3 were dominant. The component of the vocal membrane was about one order of magnitude larger than the lower mass component. The component of the upper mass m_2 was about two orders of magnitudes smaller than the vocal membrane component. For larger vocal membrane heights the components of the lower mass and the vocal membrane were of similar magnitude. The magnitude of the upper mass was significantly increased.

For small vocal membrane heights and for small tuning parameter values q , all model masses vibrated out-of-phase. At higher values of q , the phase difference between the lower and upper mass approached -120 degrees whereas the phase difference between the upper mass and the vocal membrane approached 0 degree. At larger vocal membrane heights, the phase differences as functions of the tuning parameter q were more complex. In particular, for large values of q , both phase differences approached about -30 degrees.

Increasing d_3 , only the amplitudes of the upper mass and the vocal membrane component of the third eigenmode changed significantly. This mode consisted of large components for the upper mass and the vocal membrane. The phase differences showed no changes with varying vocal membrane height. The phase difference $\varphi_2 - \varphi_1$ remained a flat function of q . Ranging between -30 degrees and 0 degree, it indicated in-phase oscillations of the lower and upper mass. The phase difference $\varphi_\theta - \varphi_2$ varied smoothly between -150 degrees and 90 degrees.

5. Vocal Membrane Model for Bats and Primates

This corresponded mainly to out-of-phase oscillations of the upper mass and the vocal membrane.

The vocal membrane height d_3 had no effect on the eigenfrequencies and relative damping times. Observe that d_3 effected only the shape of the eigenmodes, i.e. the complex amplitudes important for the phase differences between the lower and upper masses and the vocal membrane plate.

These observations suggested that d_3 was an important parameter controlling the oscillation onset behavior of the full nonlinear model. In the two-mass model, efficient energy transfer from air flow to model mass vibrations is known to depend on the phase difference between the lower and upper mass. The vocal membrane height could be a crucial geometrical control parameter for the onset of oscillations of the full nonlinear vocal membrane model.

5.3.2 Hopf bifurcation and voice onset

In this section, voice onset as a function of the geometrical and dynamical control parameters θ_0 , d_3 , and q_m was studied. Voice onset corresponds to a Hopf bifurcation where a stable fixed point becomes unstable and a limit cycle oscillation becomes stable. The fixed point was a nontrivial steady state solution of the vocal membrane model. At the Hopf bifurcation the fixed point became unstable. In terms of linear stability analysis, small oscillations became marginally stable at the bifurcation point. As follows, the frequency of these oscillations is called Hopf frequency.

In Fig. 5.16, the influence of the mass ratio q_m and the resting angle θ_0 on voice onset pressure p_s and Hopf frequency is shown. Due to $q = 1.0$, this bifurcation analysis revealed voice onset of the low frequency register of the vocal membrane model. This allowed comparison of onset behavior with the two-mass model (Steinecke and Herzel, 1995).

It was observed that the minimum onset pressure for the given parameter values was at approx. 0.9 cmH₂O. Typical threshold pressure values for the two-mass model were at about 2 cmH₂O. For small values of the resting angle θ_0 (here for $\theta_0 = 0.01, 0.1, 1.0$), the onset pressure showed little variation despite large variation of the mass ratio q_m ($1 < q_m < 5000$). As θ_0 increased, the onset pressure was shifted to significantly higher values. For very high mass ratios q_m , i.e. $m_2 \gg m_3$, the increase became negligible small.

The Hopf frequency showed little variation with q_m for small resting angles $\theta_0 = 0.01, 0.1, 1.0$. At larger values of θ_0 and small q_m , the Hopf frequency was shifted from approx. 125 Hz to approx. 140 Hz. At large values of q_m , an increasing resting angle decreased the Hopf frequency slightly.

This study showed that at small and medium values of the mass ratio q_m , the resting angle θ_0 could have a large effect on the onset pressure and Hopf frequency.

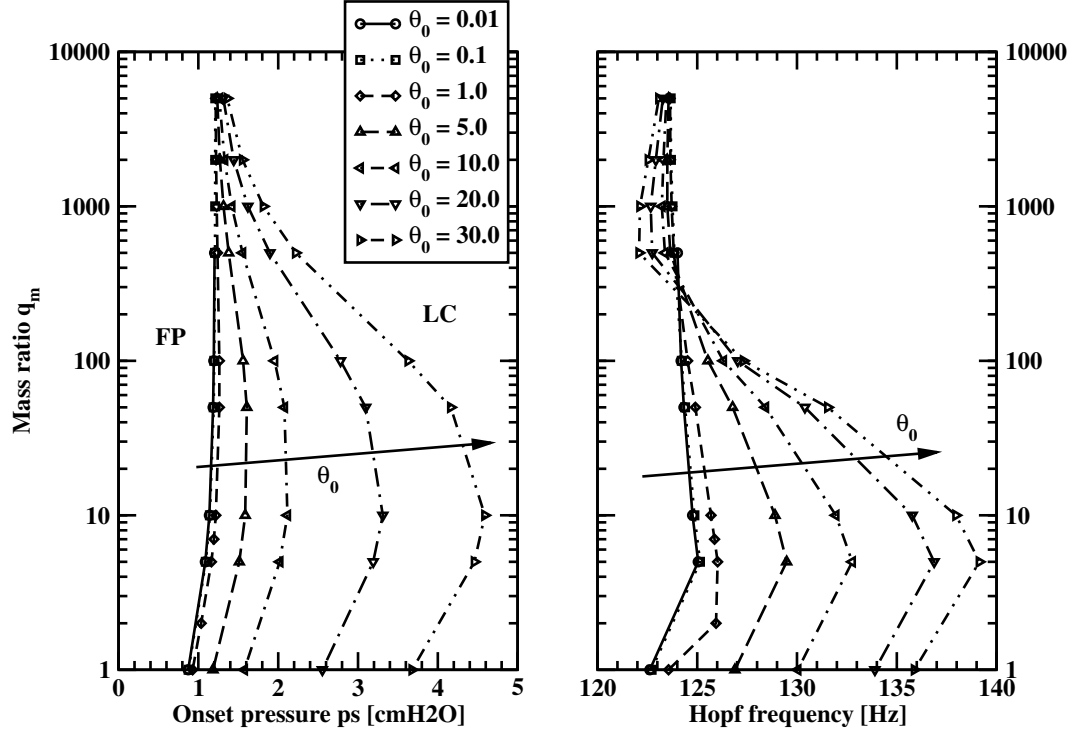


Figure 5.16: Voice onset pressure p_s and Hopf frequency at voice onset as a function of mass ratio q_m and resting angle θ_0 . At voice onset, the fixed point (FP) becomes unstable and a limit cycle oscillation (LC) becomes stable ($q = 1.0$, $d_3 = 0.05$ cm, otherwise standard parameter values are used.).

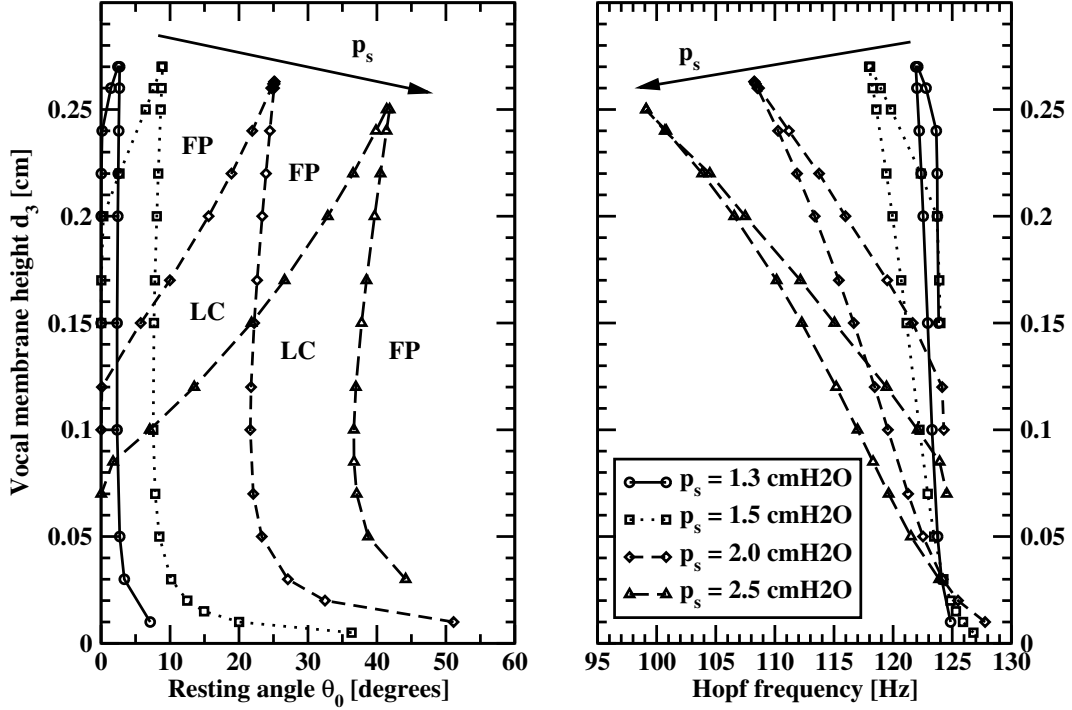


Figure 5.17: Influence of vocal membrane geometry on phonation onset. For different subglottal driving pressures p_s , phonation onset curves $d_3(\theta_0)$ and their corresponding Hopf frequencies are shown ($q_m = 500.0$, $q = 1.0$, otherwise standard parameter values are used.).

Both could be significantly increased using large values of θ_0 . At high values of q_m this influence vanished.

Fig. 5.17 shows oscillatory regions in the d_3 – θ_0 parameter subspace and their corresponding Hopf frequencies. The subglottal driving pressure p_s was kept constant along the threshold curves $d_3(\theta_0)$. As the tuning parameter was $q = 1$, this analysis showed the onset behavior of low frequency oscillations. These results could be compared to the two-mass model that served as the basis for the vocal membrane model. The mass ratio at $q_m = 500$ indicates a very lightweight vocal membrane compared to the upper vocal fold mass. The analysis for lower q_m -values showed qualitatively similar results.

The left part of Fig. 5.17 reveals the d_3 – θ_0 parameter regions where limit cycle oscillations were stable. For d_3 larger than a critical vocal membrane height (here approx. 0.27 cm), no Hopf bifurcation could be found. In general, the critical vocal membrane height was a function of the resting angle θ_0 , the mass ratio q_m and the tuning parameter q . For constant p_s the range of resting angles θ_0 with stable limit

cycle oscillations increased with decreasing d_3 . For decreasing d_3 and constant p_s the lower limit of the θ_0 -range approached zero degrees. When p_s was increased, the limiting value of d_3 for oscillations with $\theta_0 = 0$ decreased.

Fig. 5.17 could be used to find the threshold pressures for onset and offset of limit cycle oscillations for a given point in the d_3 - θ_0 -plane. For example, at $\theta_0 \approx 22$ degrees and $d_3 = 0.15$ cm, the onset pressure was $p_s = 2.0$ cmH₂O. The offset pressure, where limit cycle oscillations became unstable again, was $p_s = 2.5$ cmH₂O.

The right part of Fig. 5.17 shows the Hopf frequency corresponding to the Hopf bifurcation points in the d_3 - θ_0 plane. As discussed before, when the vocal membrane resting angle θ_0 was large, the onset pressure increased; and the corresponding Hopf frequency shifted towards smaller values. This effect was larger for larger vocal membrane heights d_3 .

In summary, small vocal membrane heights d_3 and resting angles θ_0 were found to be beneficial for a low onset pressure p_s . Such values also facilitated a large phonatory pressure range for self-sustained oscillations at low frequencies.

In Fig. 5.18, the influence of the tuning parameter q and the mass ratio q_m on phonation onset and offset and the corresponding Hopf frequencies is shown. It was observed that, for $\theta_0 = 1.0$ degree, an increasing mass ratio shifted the onset curve $p_s(q)$ only slightly to higher onset pressures. The offset pressures were shifted to smaller p_s -values. The overall area for limit cycle oscillations decreased with increasing mass ratio q_m . From the study shown before in Fig. 5.16 it can be expected that for larger θ_0 this effect is increased.

The Hopf frequencies corresponding to the onset curves $p_s(q)$ were shifted to smaller values for increasing q_m . This effect was expected to be increased with increasing resting angles θ_0 (see Fig. 5.16).

The mass ratio was found to have only a small effect on the threshold pressure at voice onset for small tuning parameter values q . In this parameter region the vocal membrane model oscillated in the low frequency register. Increasing the mass ratio decreased the maximum phonatory pressure range of approx. 7.2 cmH₂O at $q = 7$ to 5.4 cmH₂O for $q = 5.4$. The corresponding phonatory frequency range decreased from about 700 Hz to 520 Hz.

5.3.3 Bifurcation diagrams and registers

Hopf bifurcation analysis revealed a register-like structure of the oscillatory regions in parameter space. The resting angle θ_0 and the vocal membrane height d_3 influenced the onset pressure and onset frequency of the two registers. In agreement with previous work (Mergell et al., 1999) an optimal geometry (in terms of θ_0) that minimizes voice onset pressure was found.

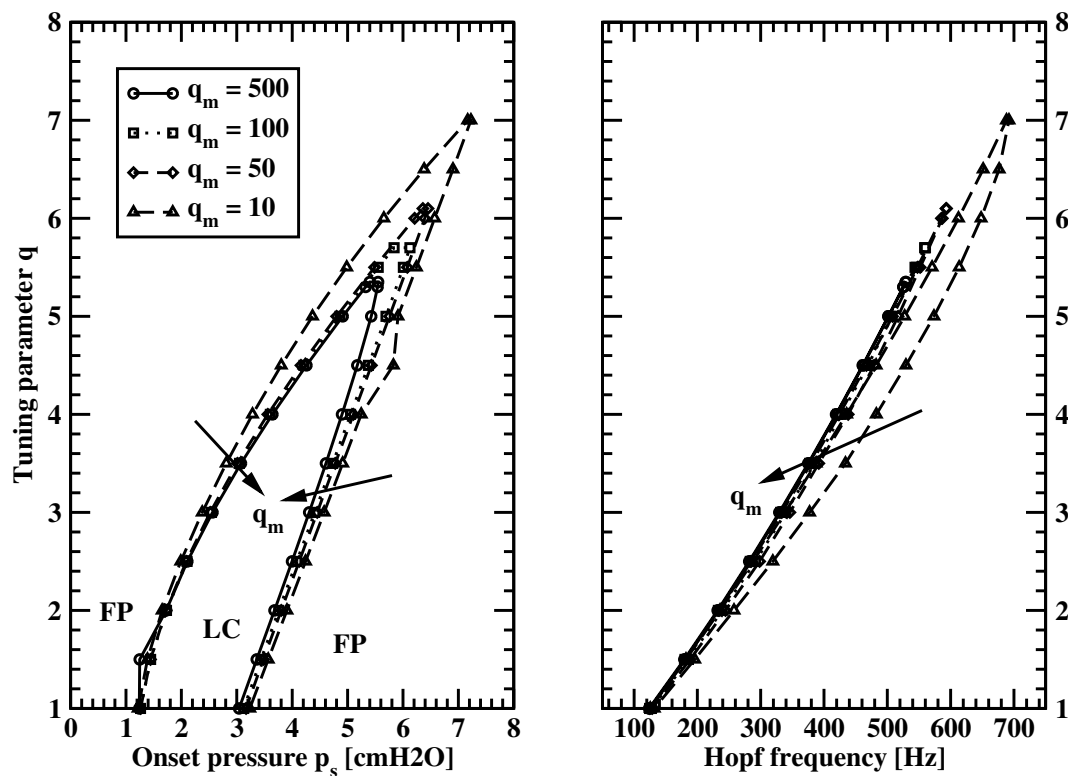


Figure 5.18: Range of low frequency oscillations varying with tuning parameter q and different mass ratios q_m . The corresponding Hopf frequencies are shown in the right part ($d_3 = 0.05$ cm, $\theta_0 = 1.0$ degree, otherwise standard parameter values are used.).

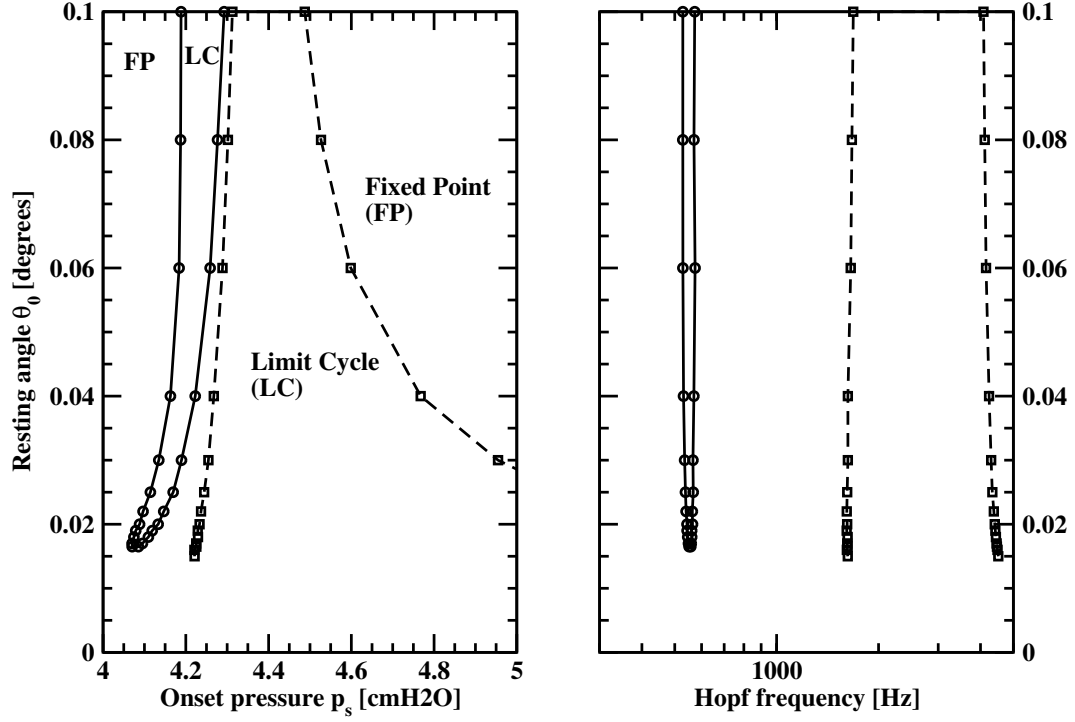


Figure 5.19: Onset pressure p_s for low and high frequency registers depending on resting angle θ_0 . The onset curve $p_s(\theta_0)$ reveals a minimum onset pressure for the low frequency register at approx. $\theta_0 = 0.015$ degrees. Note that the corresponding Hopf frequencies are shown in a semi-logarithmic plot ($q = 5.0$, $q_m = 5.0$, $d_3 = 0.1$ cm).

5. Vocal Membrane Model for Bats and Primates

Fig. 5.19 reveals regions of low and high frequency limit cycle oscillations in the θ_0 - p_s plane. For fixed θ_0 and increasing subglottal driving pressure, the fixed point became unstable and a limit cycle oscillation at low frequency became marginally stable. Further increasing p_s , the fixed point became stable again. At a second threshold pressure, another Hopf bifurcation with higher frequency occurred.

Varying the resting angle θ_0 , a minimum onset pressure was found at small θ_0 -values close to zero. In general, the minimal onset pressure depended on the vocal membrane height d_3 , the mass ratio q_m , and the tuning parameter q . At this point (here $p_s \approx 4.05$ cmH₂O, $\theta_0 \approx 0.15$ degrees) limit cycle oscillations at low frequency became marginally stable. This finding of an optimal geometry agreed with previous work on rigid reed-like vocal membranes (Mergell et al., 1999) where, for $d_3 = 0.1$ cm, a minimum of phonation onset pressure as a function of resting angle was found.

For low frequency oscillations the phonatory pressure range increased slightly with increasing resting angles. For high frequency oscillations, the pressure range was found to decrease with increasing resting angles. Both phonatory pressure ranges were small and very close to each other. Small variations of the subglottal driving pressure were sufficient to induce transitions between the low and the high frequency limit cycle oscillation.

Fig. 5.20 shows a register-like structure at voice onset. Phonation with low and high frequency limit cycle oscillations depended on the tuning parameter q . The vocal membrane height d_3 also influenced the register regions in the q - p_s parameter space.

For small d_3 and small tuning parameter values (here $0 \leq q \leq 7$) limit cycle oscillations at low frequencies were observed within a limited range of subglottal pressure ($1.2 \text{ cmH}_2\text{O} \leq p_s \leq 7.2 \text{ cmH}_2\text{O}$). At high tuning parameter values (here $q \geq 6$) high frequency oscillations were found within the same subglottal pressure range as for low frequency limit cycle oscillations. For $d_3 = 0.05$ cm both oscillatory regions overlapped. This was due to a Hopf bifurcation of the unstable fixed point in the oscillatory region at low frequencies. At the intersection of the Hopf bifurcation lines (e.g., $p_s = 6.1$ cmH₂O, $q = 6.3$) the fixed point became unstable as the growth rate of two infinitesimal small disturbances with different frequencies (here 630 Hz and 1460 Hz) were zero.

When increasing the vocal membrane height, the area for low frequency oscillations decreased. The region for high frequency oscillations was shifted to higher tuning parameter values q . Therefore, the overlap of the oscillatory regions vanished. As d_3 increased, the Hopf frequencies for the low frequency region slightly shifted to lower values. Thus, the Hopf frequencies for the high frequency region increased significantly.

For small d_3 -values, the distance between the low and the high frequency reg-

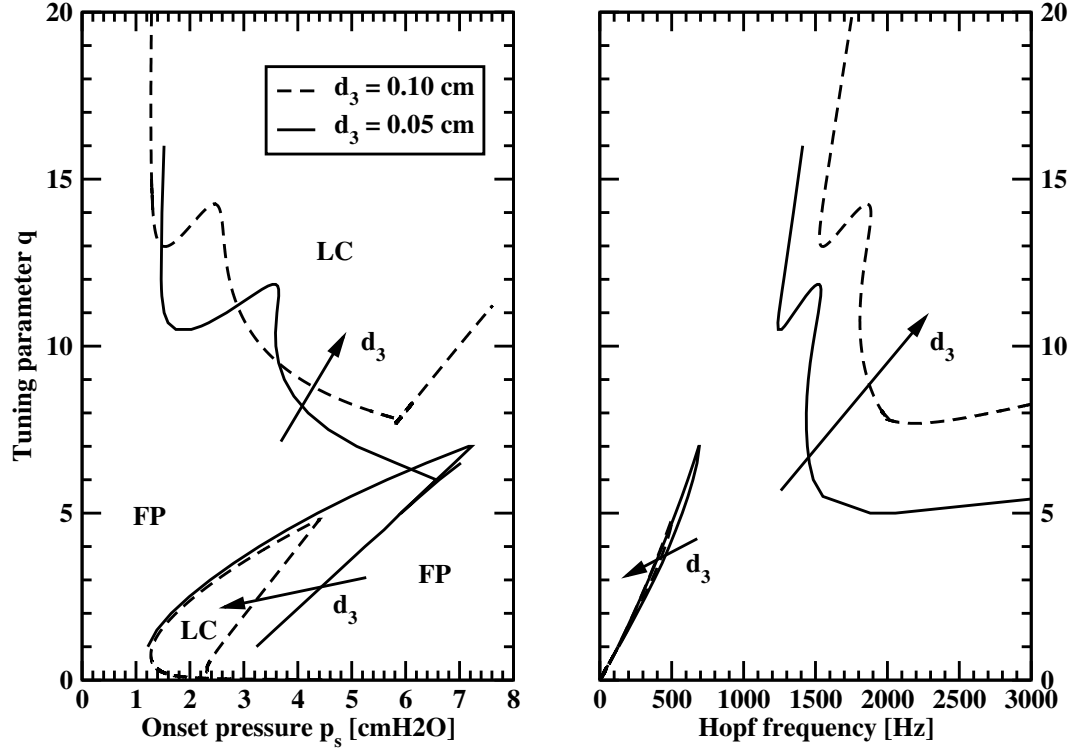


Figure 5.20: Onset pressure p_s for low and high frequency register as a function of tuning parameter q and vocal membrane height d_3 . The corresponding Hopf frequencies for the onset curves $p_s(q)$ show the register frequencies as functions of the tuning parameter q ($q_m = 10.0$, $\theta_0 = 1.0$ degree).

ister $\Delta q(p_s)$ as a function of subglottal pressure could be infinitely small. For subglottal pressures in the range $6 \text{ cmH}_2\text{O} \leq p_s \leq 7.2 \text{ cmH}_2\text{O}$, small variations of the tuning parameter q were sufficient to induce abrupt transitions between the low and the high frequency limit cycle oscillation. For larger vocal membrane heights, the register distance $\Delta q(p_s)$ had a finite positive minimum. Large variations of the tuning parameter were necessary for a register transition. For q -variations smaller than the register distance $\Delta q(p_s)$ the fixed point solution became stable again. As a consequence, phonation would be interrupted during a register transition (increase of q).

5.3.4 Simulation of complex vocalizations

The vocal membrane model was developed to mimic complex vocalizations observed in primates and echolocating bats (see Introduction of this chapter). *In vivo* experiments and recordings of squirrel monkey vocalizations involved slow variations of control parameters such as subglottal pressure and myo-elastic properties of intrinsic muscles. Values of visco-elastic properties were difficult to obtain in these experiments. *In vitro* experiments with excised larynges from squirrel monkeys allowed the control of driving air pressure values and tissue stiffness to a certain degree. Nevertheless, detailed information on the intrinsic state of excised larynges was difficult to obtain. Therefore, model parameters could not directly be derived from experiments. The vocal membrane model could be used as a “nonlinear curve fit” for complex vocalizations in primates and bats. A qualitative understanding of observed time series of primates and bats could be achieved.

Attractors of high and low frequency register

Bifurcation analysis revealed the existence of a low and high frequency register in the vocal membrane model. The dynamics of the full vocal membrane model with fixed parameter values typical for these registers was simulated. The results could be compared with *in vitro* experiments of excised squirrel monkey larynges (Brown et al., 2003). They could also shed light on the pulse generation mechanism in echolocating bats (Fattu and Suthers, 1981).

Fig. 5.21 shows a simulated time series for the vocal fold mass excursions $x_{1l}(t)$ and $x_{2l}(t)$ and the vocal membrane angle $\theta_l(t)$. The index “l” indicates the coordinates of the left model masses. Starting from standard initial conditions (see Eqs 5.31), the lower masses m_1 and m_2 performed self-sustained oscillations with a frequency of approx. 140 Hz. The tuning parameter value $q = 1$ indicated no frequency tuning of the lower two vocal fold masses, m_1 and m_2 . A phase shift between the time series x_{1l} and x_{2l} was observed, in agreement with observations

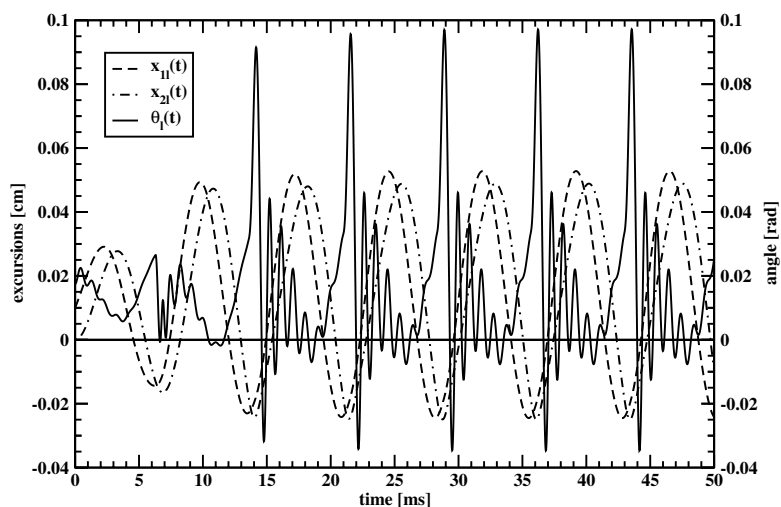


Figure 5.21: Simulated time series showing low frequency oscillation of the vocal fold masses and transient damped oscillations of the vocal membrane (The tuning parameter was set to $q = 1.0$, otherwise standard parameter values are used.).

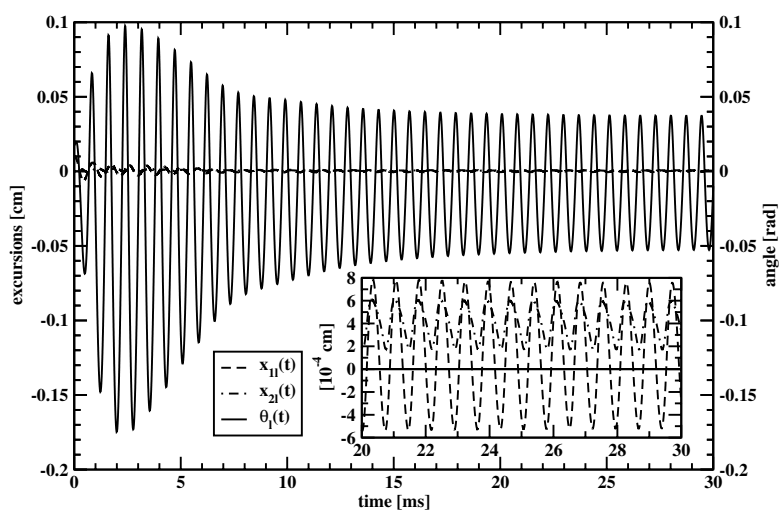


Figure 5.22: Simulated time series showing high frequency oscillation of the vocal membrane and the vocal fold masses (The tuning parameter is $q = 10.0$, otherwise standard parameter values are used.).

5. Vocal Membrane Model for Bats and Primates

on the two-mass model (Steinecke and Herzel, 1995), and the eigenmode analysis (eigenmode 3) discussed above.

The time series of the vocal membrane $\theta_l(t)$ showed large positive excursions during vocal fold contact ($x_{1l}, x_{2l} \leq -0.018$ cm). Only visco-elastic restoring forces and the forces due to dynamic coupling to the upper mass m_2 accelerated the vocal membrane mass during these time intervals. Upon glottal opening ($x_{1l}, x_{2l} \geq -0.018$ cm) transient damped oscillations of the vocal membrane interacting with the airstream were observed. The oscillation frequency of this excitable oscillator was approx. 1000 Hz.

This behavior could be associated with pulsed echolocation calls in bats. In bats the vocal folds are thought to vibrate at frequencies corresponding to low pulse rates of up to 200 Hz. Vocal membranes are supposed to perform a few hundred cycles of damped oscillations during glottal opening. In the present model only six cycles were observed. It could be expected that for a higher vocal membrane frequency, sufficiently many cycles should occur.

Observations of sustained series of pulses in experiments with excised squirrel monkey larynges have been reported by Brown et al. (2003). There, pulses with pulse rates of 8–20 Hz combined with high frequency oscillations were found. The simulation of the vocal membrane model for $q = 1$ could also mimic such experimental observations.

For a tuning parameter value of $q = 10$ for vocal fold masses, the simulated time series for excursions $x_{1l}(t)$ and $x_{2l}(t)$ and the vocal membrane angle $\theta(t)$ are shown in Fig. 5.22. The insert shows an enlarged plot of the time series $x_{1l}(t)$ and $x_{2l}(t)$. All vocal fold masses and the vocal membrane revealed self-sustained oscillations with a frequency of approx. 1400 Hz. The lower masses vibrated in phase. They oscillated out of phase related to the vocal membrane angle $\theta(t)$, i.e. all glottal areas a_1 , a_2 and a_{vm} oscillated in phase. This behavior is similar to the third eigenmode found in the eigenmode analysis before. No vocal fold contact could be observed. The oscillation amplitude of the lower mass m_1 was about twice as large as the amplitude of the upper mass m_2 . The vibration amplitude of the vocal membrane was about two orders of magnitude larger than the amplitudes of the vocal fold masses.

This behavior could resemble constant frequency (CF) echolocation calls in bats and high frequency, peeplike vocalizations in primates.

Slow variation of tuning parameter q

In high speed video experiments with brainstem stimulated squirrel monkeys rocking of the arytenoid cartilages was observed during vocalizations (Fitch et al., 2001). The length and thus the stiffness of the vocal folds were modulated sinusoidally. These modulations resulted in sinusoidal variations of the fundamental

5. Vocal Membrane Model for Bats and Primates

frequency of the primate vocalizations. The tuning parameter q of the vocal membrane model was used to mimic this behavior. It increases the restoring stiffness of the lower vocal membrane masses and decreases their vibratory masses.

In numerical experiments for the vocal membrane model, the tuning parameter q was linearly increased from $q = 1$ to $q = 10$. This range was chosen following the results of the eigenfrequency analysis. There the ratios of eigenfrequencies were shown to be mainly influenced by the tuning parameter q . Mode locking and entrained oscillations were expected for $1 \leq q \leq 10$.

Fig. 5.23 shows a spectral bifurcation diagram of the absolute vocal membrane angle $\theta_a(t)$ as the tuning parameter was increased. The resting angle $\theta_0 = 10.0$ degrees was chosen for more instabilities to occur. Fig. 5.24 shows the corresponding Poincaré section. Subsequent maxima of $\theta_a(t)$ are plotted against the tuning parameter q . The model was integrated for constant starting value $q = 1$ until stable limit cycle oscillations were observed. Then the tuning parameter was increased linearly. Thus, the fundamental frequency of the vocal membrane vibrations increased. At approx. $q = 2$, a short interval with folded limit cycle oscillations at a constant frequency was observed. At approx. $q = 3$ a period doubling bifurcation occurred. The subharmonic oscillations after the bifurcation vanished again at about $q = 3.5$. Between $q = 4$ and $q = 5$ irregular chaotic oscillations were found. Folded limit cycle oscillations reappeared and ceased at about $q = 5.8$. Between $q = 5.8$ and $q = 7$ toroidal oscillations could be found. The insert in Fig. 5.24 shows an enlarged plot of the Poincaré sections for $6 \leq q \leq 8.5$ which indicated toroidal oscillations for approx. $6 \leq q \leq 7$. At $q = 7$ a frequency jump occurred. The fundamental frequency was then approx. 1700 Hz. The spectral component at $q = 7$ and about 850 Hz indicated subharmonic oscillations. In the interval $7.2 \leq q \leq 8.3$ toroidal oscillation with a fundamental frequency of approx. 1000 Hz appeared. The distance to the sidebands in the spectrogram decreased with increasing tuning parameter q . At approx. $q = 8.3$ the fundamental frequency jumped to approx. 1700 Hz. The frequency of the limit cycle oscillation remained constant during further increases of the tuning parameter between $8.3 \leq q \leq 10$.

In the spectral bifurcation diagram (Fig. 5.23) broadened and high intensity harmonics could be observed for $2 \leq q \leq 4$. Simulations with a constant tuning parameter value $q = 1$ (Fig. 5.21) showed damped oscillations of the vocal membrane. This behavior corresponded to the formant-like structures found in the spectral bifurcation diagram where the intensity of certain harmonics was reinforced for $2 \leq q \leq 4$. In human voice formants are the acoustical resonance frequencies of the vocal tract and the subglottal trachea. Here, the vocal membrane oscillator was observed to behave in a similar way to the passive, acoustical resonators in human voice.

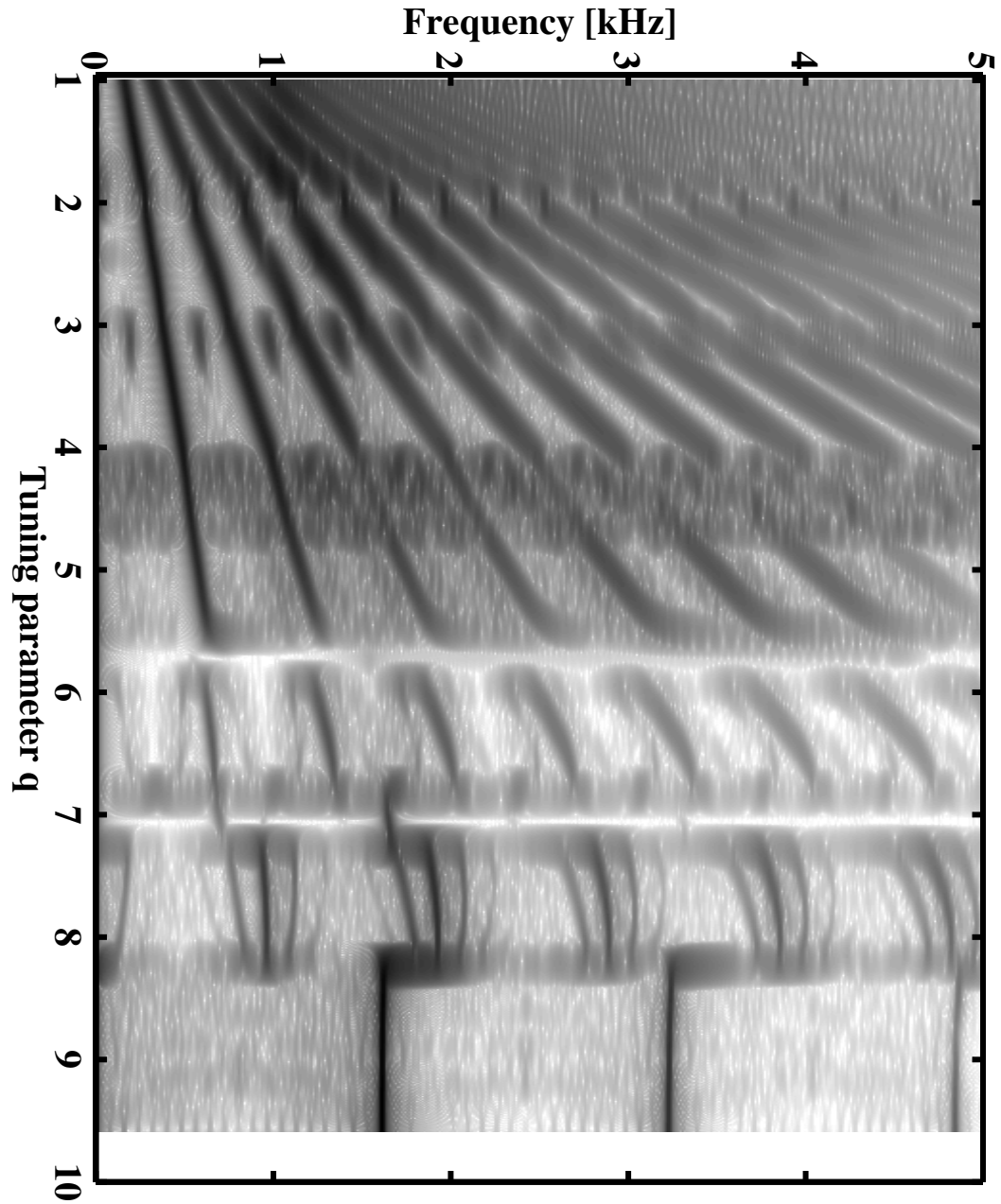


Figure 5.23: Spectral bifurcation diagram of $\theta_a(t)$ for upward gliding tuning parameter variation (linear tuning parameter variation $q(t)$, $1 \leq q \leq 10$, $p_s = 8.0$ cmH₂O, $\theta_0 = 10.0$ degrees, $q_m = 10.0$, $d_3 = 0.05$ cm, otherwise standard parameter values, simulation time: 2000 ms)

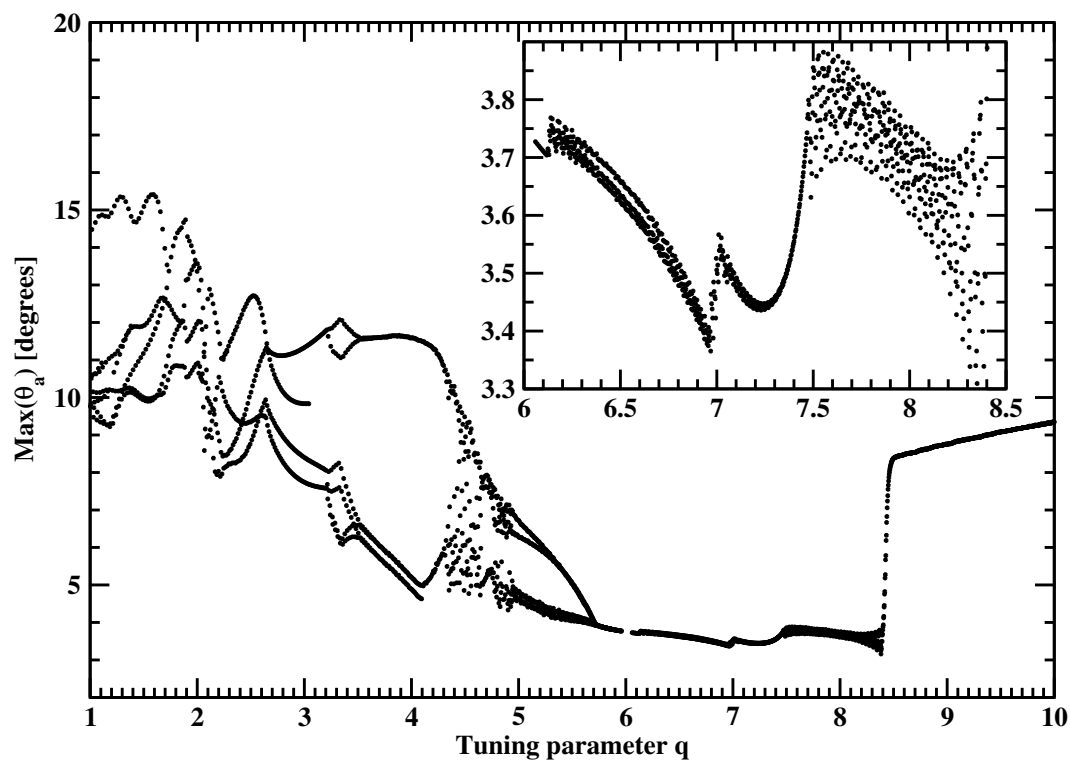


Figure 5.24: Poincare section $\dot{\theta}(t) = 0$ showing subsequent maxima of $\theta_a(t)$ for upward gliding tuning parameter variation (linear tuning parameter variation $q(t)$, $1 \leq q \leq 10$, $p_s = 8.0$ cmH₂O, $\theta_0 = 10.0$ degrees, $q_m = 10.0$, $d_3 = 0.05$ cm, otherwise standard parameter values, simulation time: 2000 ms)

Slow variation of subglottal pressure p_s

In Fig. 5.25 a spectral bifurcation diagram for linearly varying subglottal pressure p_s is shown. The pressure was slowly decreased from $p_s = 37$ cmH₂O to $p_s = 10$ cmH₂O. The model was integrated at constant pressure $p_s = 37$ cmH₂O until stable folded limit cycle oscillations with a fundamental frequency of approx. 1500 Hz were obtained. The corresponding Poincaré section in Fig. 5.26 shows subharmonic oscillations of the absolute vocal membrane angle $\theta_a(t)$ at this starting pressure. By decreasing p_s , sidebands of the fundamental frequency appeared in the spectrogram. This indicated toroidal oscillations of the vocal membrane angle. At about $p_s = 31$ cmH₂O, the onset of irregular chaotic behavior could be observed. In the chaotic region, the shading of the spectrogram showed large intensity in a narrow spectral band at about 1500 Hz. This corresponds to an unstable limit cycle oscillation of model. In the interval $29.5 \text{ cmH}_2\text{O} \leq p_s \leq 28 \text{ cmH}_2\text{O}$ a window with toroidal oscillations occurred. At about $p_s = 24$ cmH₂O another short window with limit cycle oscillation with a fundamental frequency of approx. 1500 Hz appeared. Toroidal oscillations reappeared at about $p_s = 22.5$ cmH₂O. At about $p_s = 17.5$ cmH₂O an abrupt transition to limit cycle oscillations at approx. 1500 Hz was observed.

The choice of the tuning parameter value $q = 7$ was motivated by the eigenfrequency analysis above. As discussed, q varied the ratios of eigenfrequencies of the model. The bifurcation diagram with downward gliding subglottal pressure suggested that the driving pressure controlled the degree of nonlinear interaction between the vocal membrane model modes. For $q = 7$ the distance between the eigenfrequencies was small. On the background of the theory of coupled oscillators, mode entrainment, toroidal oscillations, and chaos could be expected. The driving airstream as the nonlinear coupling induces these different behaviors.

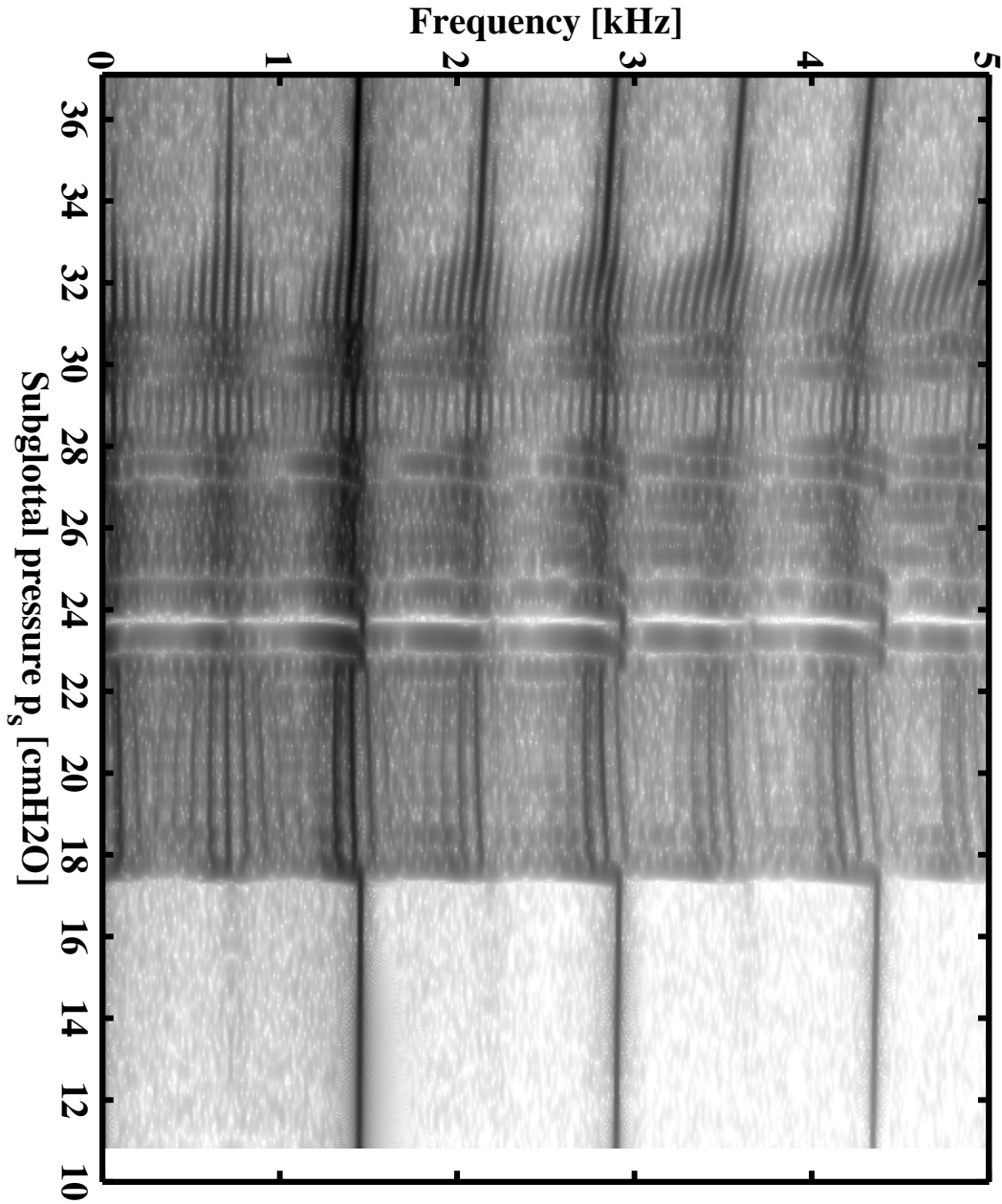


Figure 5.25: Spectral bifurcation diagram of $\theta_a(t)$ for downward gliding subglottal pressure variation (linear pressure variation $p_s(t)$, $37.0 \text{ cmH}_2\text{O} \geq p_s \geq 10.0 \text{ cmH}_2\text{O}$, $q = 7$, $\theta_0 = 1.0$ degree, $q_m = 10.0$, $d_3 = 0.05 \text{ cm}$, otherwise standard parameter values, simulation time: 3000 ms)

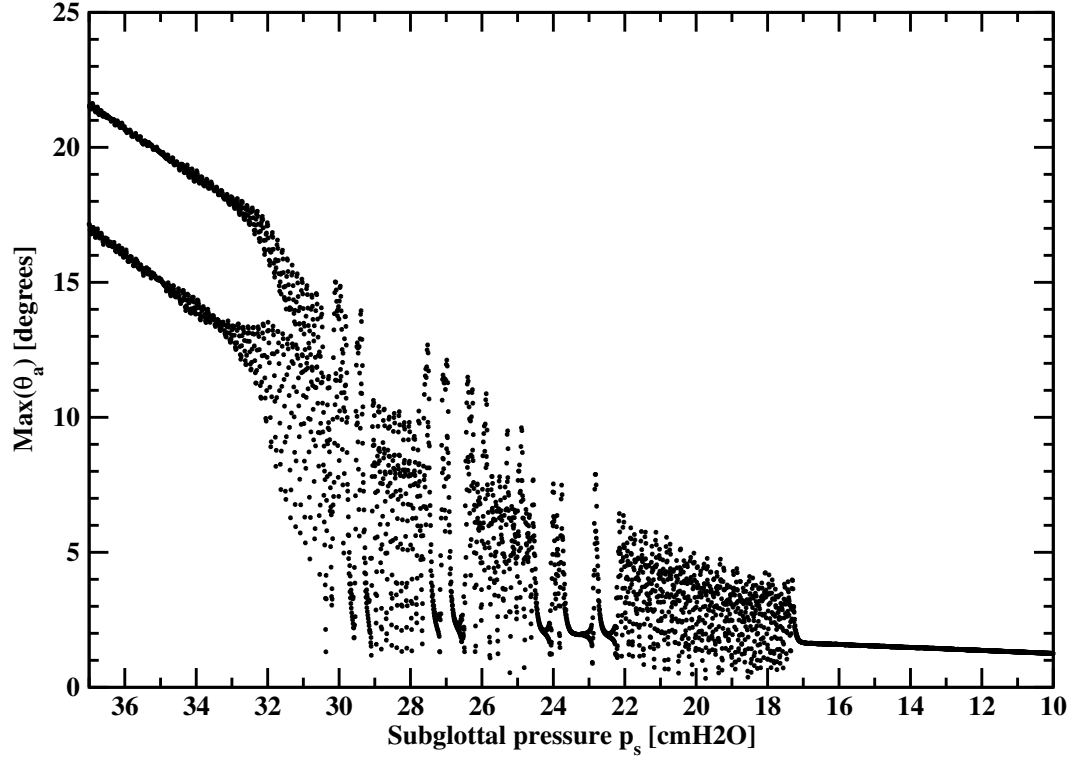


Figure 5.26: Poincare section $\dot{\theta}(t) = 0$ showing subsequent maxima of $\theta_a(t)$ for downward gliding subglottal pressure variation (linear pressure variation $p_s(t)$, $37.0 \text{ cmH}_2\text{O} \geq p_s \geq 10.0 \text{ cmH}_2\text{O}$, $q = 7$, $\theta_0 = 1.0$ degree, $q_m = 10.0$, $d_3 = 0.05 \text{ cm}$, otherwise standard parameter values, simulation time: 3000 ms)

5.4 Discussion

In this chapter, a phenomenological dynamical model for vocal membranes, a widespread variation in nonhuman mammalian larynges, was developed. It was based on acoustical data of echolocation calls in bats and on the diverse sound repertoire of squirrel monkeys. Anatomical details for nonhuman larynges in bats and primates were incorporated into the model only qualitatively. Neurophysiological data for squirrel monkeys and bats evidenced the need for a biomechanical-aerodynamical model for vocal membrane systems in nonhuman mammals. Neural control of the internal state of the larynx could be modeled by slowly varying system parameters. Indirect *in vivo* and *in vitro* observations in primates showed complex vibratory patterns of larynges with vocal membranes. Recent *in vivo* high speed recordings supported the dynamic significance of vocal membranes in squirrel monkeys. These studies support the model hypothesis of interacting oscillations of vocal fold and vocal membrane tissue in nonhuman primates.

The mathematical description of the dynamics of the model was derived from a simplified mechanical model with a small number of mechanical degrees of freedom. It extended previous models on human voice production and a model with static vocal membranes. The aerodynamic description was based on a modified potential flow law where detached air flow was incorporated with a simple rule. Due to rigid mechanical coupling of the vocal membranes to the vocal folds, new dynamic coupling terms appeared in the well-known two-mass model equations. This coupling was due to centrifugal forces of the vocal membrane on the vocal fold and due to conservation of linear and angular momentum.

From a methodological viewpoint, the extended model is required to include as much experimental knowledge as possible by using as many new variables and new control parameters as necessary. Five biologically relevant control parameters were introduced. First, a frequency tuning parameter for the vocal folds corresponded to activation of intrinsic laryngeal muscles and the resulting stiffness variation of the vocal fold tissue. Next, the oscillation frequency of the vocal membranes was independently fixed at a value an order of magnitude higher than the typical oscillation frequency of the vocal folds in humans. Then, the mass ratio of the dynamic mass of the vocal fold and the vocal membrane was introduced as a control parameter due to scaling of the vocal membrane equations. Finally, the geometric parameters for vocal membrane height and the resting angle of the vocal membrane during prephonatory standstill completed the set of biologically important control parameters.

The main focus of the analysis of the vocal membrane model was on dynamic effects of vibratory vocal membranes as additional oscillators atop the vocal folds. Linear eigenmode analysis revealed the dependence of eigenfrequencies and eigenmodes on the geometric specifications of the vocal membranes and on the ratio of

vocal fold frequencies and vocal membrane frequency. Phonation onset following prephonatory standstill was studied in terms of Hopf bifurcations of the model from prephonatory fixed points. The influence of the geometric parameters, the mass ratio, and the frequency ratio between vocal folds and vocal membranes on the onset pressure was investigated. Simulations with slowly varying subglottal driving pressure values and frequency tuning showed that the model could reproduce harmonic vocalizations, subharmonic sound patterns, toroidal oscillations and irregular chaotic calls in primates and bats. Abrupt transitions between different vibratory behaviors (subharmonic oscillation, torus, chaos) and oscillatory registers in the model were agreed qualitatively with observations in bats and primates.

5.4.1 Registers in larynges with vocal membranes

The most significant finding of the bifurcation analysis and of simulations of the vocal membrane model is the existence of two different voice registers at voice onset. The onset pressure as a function of the frequency tuning parameter q of the vocal folds was obtained with Hopf bifurcation analysis. Two different onset behaviors were observed. At small values of q , phonation onset occurred at subglottal pressures comparable to values found in the previous static vocal membrane model (Mergell et al., 1999). In this low register, the frequency at onset corresponded to the frequency of the tuned two-mass model. It smoothly increased with increasing q . At higher values of q , the oscillation frequency at phonation onset was on the order of the prescribed vocal membrane frequency (here 1000 Hz). The minimal onset pressure for this high register was comparable to the onset pressure of the low register (here about 1.2 cmH₂O). In general, the separation of the register boundaries $\Delta q(p_s)$ was a function of the vocal membrane height, the mass ratio, and the resting angle of the vocal membranes.

It was shown that the two register boundaries could overlap. Then, infinitesimally small changes of the tuning parameter at constant driving pressure could change the oscillatory behavior from the low register to the high register. A frequency jump (here about 1000 Hz) could be the consequence. At the crossings of the register boundaries, the vocal membrane model could oscillate at low and high frequency simulatenously. On the background of the theory of coupled oscillators, entrained oscillation, toroidal behavior and chaos can be expected in the vicinity of these points. This asymptotic dynamics depends on the coupling, e.g., due to the aerodynamic driving, and on the path in parameter space that was used to reach these points.

5.4.2 Diversity of vocal membrane dynamics

Another important feature of the presented vocal membrane model is the diversity of dynamic behaviors induced by the two interacting oscillators – the vocal folds and the vocal membranes. Unlike human speech, communication among nonhuman mammals (here in particular primates and bats) is based mostly on the diverse sound repertoire produced in the larynx. Information coding in mammals relies on complex phonatory patterns in contrast to complex articulation patterns of the vocal tract in human speech.

The vocal membrane is an oscillator on top of the vocal folds and adds another oscillation to the set of coupled oscillatory modes of the vocal folds. It has a significantly higher frequency than the typical excited modal frequencies of the vocal folds. Mode entrainment and synchronization phenomena are governed by the frequency ratio between these oscillators. The introduced tuning parameter q of the vocal folds is related to this ratio. It can be associated with the frequency ratio known from the generic circle map for coupled oscillators. Linear eigenmode analysis showed the influence of the vocal membrane geometry (height d_3 , resting angle θ_0 , mass ratio q_m) on the tuning parameter in terms of entrainment regions of the coupled oscillators. They could be related to the Arnold tongues found in the circle map. Note that simulations also showed the existence of coexisting (folded) limit cycles oscillations at fixed points and within the Hopf bifurcation region of the register diagram (Fig. 5.20). Thus, coexisting attractors further increased the dynamic richness of the vocal membrane model.

The diversity of oscillatory behavior found in simulations of the vocal membrane model with gliding parameter variations (here the subglottal pressure p_s and the tuning parameter q) is thought to be induced by the Arnold tongue-like structure of the phase space of the vocal membrane model. The subglottal aerodynamic driving force can be related to the nonlinear coupling in the circle map. Simulations with gliding subglottal pressure variations showed subharmonic oscillations, tori, register jumps and irregular chaotic oscillations. Similar findings were obtained in numerical simulations for tuning parameter variations.

In contrast to the static vocal membrane model (Mergell et al., 1999), these voice instabilities occurred without additional left-right asymmetry. The static model did not show instabilities, even for high driving pressures, unless left and right vocal folds were made significantly different in mass and stiffness. The static model also lacked toroidal oscillations and frequency jumps. Due to the dynamic vocal membrane, these instabilities are an inherent feature of the new vocal membrane model. They allow understanding of voice instabilities frequent in primate vocalizations and in echolocation calls in bats.

Simulations at fixed parameter values corresponding to the low register showed low frequency oscillations of the vocal folds combined with high frequency damped

5. Vocal Membrane Model for Bats and Primates

oscillations of the attached vocal membrane. This behavior mimics pulsed echolocation calls in FM bats (echolocating bats using downward sweeping frequency modulated calls). During prey capture FM bats abruptly increase the pulse repetition rate of ultrasound FM calls to rates up to 200 Hz. Modeling shows that such terminal feeding buzzes in bats might be accomplished by low frequency oscillations of the vocal folds. The vocal folds could provide the pulse modulation of ultrasound vocalizations generated by high frequency oscillations of the vocal membranes. This model of terminal feeding buzzes in bats might require only simple neural control. Bats would have to induce only the transition between single ultrasound calls at low repetition rates to dynamic modulations of the high frequency vocal membrane oscillations by the self-sustained vibrating vocal folds. This control could be accomplished by stiffness changes within the vocal folds.

Model simulations for parameter values associated with the high register could resemble echolocation calls in CF bats (bats using constant frequency calls) and trill calls (peep-like calls) in squirrel monkeys. Peep-like calls are purely high pitched vocalizations with a sinusoidal modulation of the fundamental frequency. As shown in the Introduction of this chapter, peep calls could abruptly bifurcate to low frequency vocalizations or irregular chaotic oscillations. This diversity is captured qualitatively with the presented dynamical vocal membrane model.

The results of gliding tuning parameter simulations suggest that vocal membranes could be regarded as passive filters. In the low register the vocal membranes perform damped oscillations at high frequency. Acoustically, they could be similar to formants as they amplify and attenuate oscillation frequencies of the primary voice source (the vocal folds). This pseudo formant induced by the vocal membranes could broadcast caller identity in primate or bat communication.

5.4.3 Adaptability and robustness of vocal membrane systems

The final finding of the present analysis is the adaptability and flexibility of vocal membrane systems. Nonhuman mammals possess an increased dimension of phonatory parameter space due to their vocal membranes.

An important physiological constraint is phonation onset pressure. It is related to the efficiency of energy transfer from the driving air flow to vocal fold and vocal membrane oscillations (Mergell et al., 1999). Hopf bifurcation analysis of the vocal membrane model showed the influence of the vocal membrane parameters on phonation onset. The low register results showed that very small and very large mass ratios in combination with vocal membrane resting angles close to zero are beneficial for a low phonation onset pressure. For medium mass ratios large resting angle increased the onset pressure the most. Additionally, high mass ratios

decreased the phonatory range of large tuning parameter values.

In terms of the Hopf bifurcation from the prephonatory fixed point there is a critical vocal membrane height d_3 for which phonation onset could be found. This critical value depended on the resting angle, the mass ratio, and the tuning parameter value. For subcritical vocal membrane heights the phonatory pressure range increases with decreasing d_3 . Furthermore, the phonatory range in terms of the resting angle is increased. This implies that small vocal membrane heights and small vocal membrane resting angles are beneficial for low phonation onset pressure. In this case the mass ratio has only a small effect on phonation onset. Furthermore, if d_3 and the resting angle are small, neural control of the subglottal pressure is not required to be very accurate as the phonatory pressure range is large.

For very small resting angles the onset pressure could be minimized. This finding is similar to results in the static vocal membrane model (Mergell et al., 1999). It also agrees with work on phonation onset in the two-mass model, where the parallel glottis was found to be best for minimal onset pressure (Lucero, 1993; Titze, 1988). A parallel glottis is similar to nearly vertical vocal membranes (zero resting angle).

The particular design of the vocal membranes in bats and primates is a trade-off between increased interaction surface of the vibratory tissue with the driving air flow and increased dynamic interaction of the vocal membrane with the vocal folds. The dynamic coupling of the vocal membranes to the vocal folds is essentially nonlinear. As shown in the Hopf bifurcation analysis, this also increases the complexity of the phase space of a vocal membrane system. Neural controllability limits the complexity of dynamic patterns and their associated basins of attraction in nature. For example, subglottal pressure ranges and tissue stiffnesses are limited, and neurally controlled variations of internal glottal configuration are not infinitesimally small. The design of the vocal membranes in terms of vocal membrane height, resting angle, mass ratio, and tissue stiffness should aim for a high degree of robustness, controllability and diversity of dynamic patterns.

5.5 Outlook

In this chapter a vocal membrane model with parameters adopted from human phonation was studied. The focus was on a qualitative understanding of the dynamic interaction of vocal membranes with vocal folds. Quantitative results could be obtained on the geometrical design specifications of vocal membranes.

A detailed “nonlinear fit” of vocalizations of bats and primates remains an open task. To date detailed anatomical data for nonhuman primate larynges and for bats are missing. Due to technical difficulties, time-resolved direct observations

5. Vocal Membrane Model for Bats and Primates

of oscillating vocal membranes are very challenging. In future they could help modeling particular calls in a quantitative way. Thus, the close cooperation of modeling and experiments on nonhuman primates with vocal membranes should continue.

Bibliography

- F. Alipour and R. C. Scherer. Effects of oscillation of a mechanical hemilarynx model on mean transglottal pressures and flows. *J. Acoust. Soc. Am.*, 110:1562–1569, 2001.
- F. Alipour-Haghighi and I. R. Titze. Simulation of particle trajectories of vocal fold tissue during phonation. In I. R. Titze and R. C. Scherer, editors, *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, pages 183–190. Denver Center for the Performing Arts Denver, CO, 1985.
- F. Alipour-Haghighi and I. R. Titze. Elastic models of vocal fold tissues. *J. Acoust. Soc. Am.*, 90:1326–1331, 1991.
- F. Alipour-Haghighi, I. R. Titze, and A. L. Perlman. Tetanic contraction in vocal fold muscle. *J. Speech Hear. Res.*, 32:226–231, 1989.
- F. Alipour-Haghighi, C. Fan, and R. C. Scherer. A numerical simulation of laryngeal flow in a forced-oscillation glottal model. *Comp. Speech Language.*, 10:75–93, 1996.
- F. Alipour-Haghighi, D. A. Berry, and I. R. Titze. A finite element model of vocal fold vibration. *J. Acoust. Soc. Am.*, 108:3003–3012, 2000.
- I. Anhalt. *Alternative voices*. University of Toronto Press, 1984.
- V. I. Arnold. Small denominators, I. Mappings of the circumference onto itself. *Trans. of the Am. Math. Soc.*, 42:213, 1965.
- N. Aubry, R. Guyonnet, and R. Lima. Spatiotemporal analysis of complex signals: theory and application. *J. Stat. Phys.*, 64:683–739, 1991.
- S. F. Austin and I. R. Titze. The effect of subglottal resonance upon vocal fold vibration. *J. Voice*, 11:391–402, 1997.
- J.-A. Bachorowski, M. J. Smoski, and M. J. Owren. The acoustic features of human laughter. *J. Acoust. Soc. Am.*, 110:1581–1597, 2001.

BIBLIOGRAPHY

- T. Baer, J. C. Gore, L. C. Gracco, and P. W. Nye. Analysis of vocal tract shape and dimensions using magnetic resonance imaging. *J. Acoust. Soc. Am.*, 90:799–828, 1991.
- B. M. Barnett. Aspects of vocal multiphonics. Master’s thesis, University of California-San Diego, 1972.
- B. Bartolozzi. *Metodo per Oboe*. Edizioni Suvini Zerboni, Milano, 1969.
- A. Behrman. Global and local dimensions of vocal dynamics. *J. Acoust. Soc. Am.*, 105:432–434, 1999.
- A. Behrman and C. J. Agresti. Microphone and electroglottographic data from dysphonic patients: type 1, 2 and 3 signals. *J. Voice*, 12:249–260, 1998.
- A. Behrman and R. J. Baken. Correlation dimension of electroglottographic data from healthy and pathologic subjects. *J. Acoust. Soc. Am.*, 102:2371–2379, 1997.
- P. Bergé, Y. Pomeau, and C. Vidal. *Order within Chaos*. Hermann and Wiley & Sons, 1984.
- G. S. Berke, D. C. Green, M. E. Smith, D. P. Arnstein, V. Honrubia, M. Natividad, and W. A. Conrad. Experimental evidence in the in vivo canine for the collapsible tube model in phonation. *J. Acoust. Soc. Am.*, 89:1358–1363, 1991.
- D. A. Berry. Mechanisms of modal and nonmodal phonation. *J. Phonetics*, 29:431–450, 2001.
- D. A. Berry, H. Herzel, I. R. Titze, and K. Krischer. Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions. *J. Acoust. Soc. Am.*, 95:3595–3604, 1994.
- D. A. Berry, H. Herzel, I. R. Titze, and B. H. Story. Bifurcations in excised larynx experiments. *J. Voice*, 10:129–138, 1996.
- D. A. Berry, D. W. Montequin, and N. Tayama. High-speed digital imaging of the medial surface of the vocal folds. *J. Acoust. Soc. Am.*, 110:2539–2547, 2001.
- S. Bielamowicz, G. S. Berke, J. Kreiman, and B. R. Gerratt. Exit jet particle velocity in the in vivo canine laryngeal model with variable nerve stimulation. *J. Voice*, 13:153–160, 1999.
- G. Bijma. *Five Voices*. Compact Disc, Intakt, Zürich, 1989.

BIBLIOGRAPHY

- M. Blomgren, Y. Chen, M. L. Ng, and H. R. Gilbert. Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *J. Acoust. Soc. Am.*, 103:2649–2658, 1998.
- J. Blonk. *Vocalor*. Compact Disc, Staalplaat, Amsterdam, 1997.
- H. Bloss, C. Backert, and A. Raguse. CAMSYS high speed camera system. *Fraunhofer Gesellschaft IIS, Erlangen, Germany*, 1993.
- S. Boinski and C. L. Mitchell. Wild squirrel monkey (*Saimiri sciureus*) "caregiver" calls: contexts and acoustic structure. *Am. J. Primat.*, 35:129–137, 1995.
- S. Boinski and C. L. Mitchell. Chuck vocalizations of wild female squirrel monkeys (*Saimiri sciureus*) contain information on caller identity and foraging activity. *Int. J. Primatol.*, 18:975–993, 1997.
- K. S. Breuer and L. Sirovich. The use of the Karhunen-Loève procedure for the calculation of linear eigenfunctions. *J. Comp. Phys.*, 96:277–296, 1991.
- A. R. Britton and G. Jones. Echolocation behaviour and prey-capture success in foraging bats: laboratory and field experiments on *Myotis daubentonii*. *J. Exp. Biol.*, 202:1793–1801, 1999.
- C. H. Brown and M. P. Cannito. Modes of vocal variation in Syke's monkey (*Cercopithecus albogularis*) squeals. *Journal of Comparative Psychology*, 109:398–415, 1995.
- C. H. Brown, F. Alipour, D. A. Berry, and D. Montequin. Laryngeal biomechanics and vocal communication in the squirrel monkey (*Saimiri boliviensis*). *J. Acoust. Soc. Am.*, 113:2114–2126, 2003.
- J. A. Castelijns, M. W. van den Brekel, V. A. Niekopp, and G. B. Snow. Imaging of the larynx. *Neuroimaging Clin. N. Am.*, 6:401–415, 1996.
- J. Chadabe. *Electric sound*. Upper Saddle River, Prentice Hall, 1997.
- R. W. Chan. Estimation of viscoelastic shear properties of vocal-fold tissues based on time-temperature superposition. *J. Acoust. Soc. Am.*, 110:1548–1561, 2001.
- R. W. Chan and N. Tayama. Biomechanical effects of hydration in vocal fold tissues. *Otolaryngol. Head Neck Surg.*, 126:528–537, 2002.
- R. W. Chan and I. R. Titze. Viscoelastic shear properties of human vocal fold mucosa: theoretical characterization based on constitutive modeling. *J. Acoust. Soc. Am.*, 107:565–580, 2000.

BIBLIOGRAPHY

- R. W. Chan and I. R. Titze. Viscoelastic shear properties of human vocal fold mucosa: measurement methodology and empirical results. *J. Acoust. Soc. Am.*, 106:2008–2021, 1999.
- R. W. Chan, I. R. Titze, and M. R. Titze. Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa. *J. Acoust. Soc. Am.*, 101:3722–3727, 1997.
- A. M. Chase. Aspects involving the performance of contemporary vocal music. Master’s thesis, University of California-San Diego, 1975.
- E. M. Clark. Emphasizing the articulatory and timbral aspects of vocal production in vocal composition. D.M.A. Thesis, University of Illinois, 1985.
- D. S. Cooper and I. R. Titze. Generation and dissipation of heat in vocal fold tissue. *J Speech Hear Res*, 28:207–215, 1985.
- B. Cranen and L. Boves. Pressure measurements during speech production using semiconductor miniature pressure transducers. *J. Acoust. Soc. Am.*, 77:1543–1551, 1985.
- B. Cranen and L. Boves. On subglottal formant analysis. *J. Acoust. Soc. Am.*, 81:734–746, 1987.
- B. Cranen and L. Boves. On the measurements of glottal flow. *J. Acoust. Soc. Am.*, 84:888–900, 1988.
- J. Dang and K. Honda. Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation. *J. Acoust. Soc. Am.*, 100:3374–3383, 1996.
- J. Dang and K. Honda. Acoustic characteristics of the piriform fossa in models and humans. *J. Acoust. Soc. Am.*, 101:456–465, 1997.
- G. de Krom. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *J. Speech Hear. Res.*, 36:224–266, 1993.
- M. de Oliveira Rosa, J. C. Pereira, M. Grellet, and A. Alwan. A contribution to simulating a three-dimensional larynx model using the finite element method. *J. Acoust. Soc. Am.*, 114:2893–2905, 2003.
- M. P. de Vries, H. K. Schutte, A. E. P. Veldman, and G. J. Verkerke. Glottal flow through a two-mass model: Comparison of navier-stokes solutions with simplified models. *J. Acoust. Soc. Am.*, 111:1847–1853, 2002.

BIBLIOGRAPHY

- L. Dolansky and P. Tjernlund. On certain irregularities of voiced–speech waveforms. *IEEE Transactions*, AU-16:51–56, 1968.
- M. Döllinger, U. Hoppe, F. Hettlich, J. Lohscheller, S. Schuberth, and U. Eysholdt. Vibration parameter extraction from endoscopic image series of the vocal folds. *IEEE Trans. Biomed. Eng.*, 49:773–781, 2002.
- C. Drioli and F. Avanzini. Hybrid parametric-physiological glottal modelling with application to voice quality assessment. *Med. Eng. Phys.*, 24:453–460, 2002.
- W. Ebeling, H. Engel, and H. Herzel. *Selbstorganisation in der Zeit*. Akademie-Verlag, Berlin, 1990.
- M. E. Edgerton. The 21st century voice. Submitted book manuscript, 2002.
- M. E. Edgerton, A. Khidr, and D. Bless. Multiple sound sources of the vocal tract (an analysis of [imitated Tibetan] chant. *National Center for Voice and Speech Status and Progress report*, 13:131–140, 1999.
- M. E. Edgerton, J. Neubauer, and H. Herzel. The influence of nonlinear dynamics and the scaling of multidimensional parameter spaces in instrumental, vocal and electronic composition. In *Proceedings of 4th Generative Art Conference*, Politecnico di Milano University, Milan, Italy, 2001. Generative Design Lab DIAP.
- M. E. Edgerton, J. Neubauer, and H. Herzel. Using nonlinear phenomena in contemporary musical composition and performance. *Perspectives of New Music*, 41:30–65, 2003.
- H. R. Erwin, W. W. Wilson, and C. F. Moss. A computational sensorimotor model of bat echolocation. *J. Acoust. Soc. Am.*, 110:1176–1187, 2001.
- U. Eysholdt, M. Tigges, and U. Pröschel. Direct evaluation of high speed recordings of vocal fold vibrations. In *Procs. of The 3rd International Symposium on Phonosurgery, Kyoto, Japan*, 1994.
- U. Eysholdt, M. Tigges, T. Wittenberg, and U. Pröschel. Direct evaluation of high-speed recordings of vocal fold vibrations. *Folia Phoniatr. Logop.*, 48:163–170, 1996.
- A. Facchini, S. Bastianoni, N. Marchettini, and M. Rustici. Characterization of chaotic dynamics in the vocalization of *Cervus elaphus corsicanus*. *J. Acoust. Soc. Am.*, 114:3040–3043, 2003.
- D. W. Farnsworth. High-speed motion pictures of the human vocal folds. *Bell Lab. Rec.*, 18:203–208, 1940.

BIBLIOGRAPHY

- J. M. Fattu and R. A. Suthers. Subglottic pressure and the control of phonation by the echolocating bat, *Eptesicus*. *J. Comp. Physiol.*, 143:465–475, 1981.
- M. S. Fee. Measurement of the linear and nonlinear mechanical properties of the oscine syrinx: Implications for function. *J. Comp. Physiol. A*, 188:829–839, 2002.
- M. S. Fee, B. Shraiman, B. Pesaran, and P. P. Mitra. The role of nonlinear dynamics of the syrinx in vocalizations of a songbird. *Nature*, 395:67–71, 1998.
- T. Fitch, J. Neubauer, and H. Herzel. Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. *Anim. Behav.*, 63:407–418, 2002.
- W. T. Fitch and J. Giedd. Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J. Acoust. Soc. Am.*, 106:1511–1522, 1999.
- W. T. Fitch and M. D. Hauser. Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on "honest" advertisement. *American Journal of Primatology*, 37:191–219, 1995.
- W. T. Fitch, O. N. Larsen, U. H. L. Häusler, U. Jürgens, B. B. Andersen, and J. Neubauer. Understanding primate vocal virtuosity: High-speed video endoscopy of vocalizing squirrel monkeys. in preparation, 2001.
- J. L. Flanagan. *Speech Analysis, Synthesis, and Perception*. Springer-Verlag, Berlin, 1965.
- J. L. Flanagan and L. Cherry. Excitation of vocal-tract synthesizers. *J. Acoust. Soc. Am.*, 45:1607–1617, 1969.
- J. L. Flanagan and L. L. Landgraf. Self-oscillating source for vocal tract synthesizers. *IEEE Trans. on Audio and Electroacoust.*, AU-16:57–64, 1968.
- N. H. Fletcher. Bird song – a quantitative model. *J. theor. Biol.*, 135:455–481, 1988.
- N. H. Fletcher and A. Tarnopolsky. Acoustics of the avian vocal tract. *J. Acoust. Soc. Am.*, 105:35–49, 1999.
- L. Fuks, B. Hammarberg, and J. Sundberg. A self-sustained vocal-ventricular phonation mode: acoustical, aerodynamic and glottographic evidences. *TMH-QPSR*, 3:49–59, 1998.

BIBLIOGRAPHY

- K. Funakoshi, A. Zubaid, and S. Matsumura. Regular pulse emission in some megachiropteran bats. *Zoological-Science-Tokyo*, 12:503–505, 1995.
- Y. C. Fung. *Biomechanics. Mechanical Properties of Living Tissues*. Springer, New York, 2nd edition, 1993.
- T. Gardner, G. Cecchi, and M. Magnasco. Simple motor gestures for birdsongs. *Phys. Rev. Lett.*, 87:208101, 2001.
- J. P. Gautier. Etude morphologique et fonctionnelle des annexes extra-larynges des cercopithecinae; liaison avec les cris d’espacement [Morphological and functional study of the extralaryngeal vocal sac of *Cercopithecinae*; connection with spacing calls]. *Biol. Gabonica*, 7:230–267, 1971.
- B. R. Gerratt and J. Kreiman. Theoretical and methodological development in the study of pathological voice quality. *J. Phonetics*, 28:335–342, 2000.
- B. R. Gerratt, K. Precoda, D. G. Hanson, and G. S. Berke. Source characteristics of diplophonia. unpublished manuscript, 1984.
- B. R. Gerratt, K. Precoda, and D. G. Hanson. Diplophonia: Features in the time domain. In *Annual convention of the American Speech-Language-Hearing Association*, New Orleans, 1987. American Speech-Language-Hearing Association.
- V. Gibiat and M. Castellengo. Period doubling occurrences in wind instruments musical performance. *Acustica Acta acustica*, 86:746–754, 2000.
- A. Giovanni, M. Ouaknine, B. Guelfucci, P. Yu, M. Zanaret, and J. Triglia. Non-linear behavior of vocal fold vibration: the role of coupling between the vocal folds. *J. Voice*, 13:465–476, 1999a.
- A. Giovanni, M. Ouaknine, and J. Triglia. Determination of largest lyapunov exponents of vocal signal: application to unilateral laryngeal paralysis. *J. Voice*, 13:341–354, 1999b.
- L. Glass and M. Mackey. *From Clocks to Chaos*. Princeton University Press, 1988.
- C. Gottwald. *”Hallelujah” und die Theorie des kommunikativen Handelns*. Klett-Cotta, Stuttgart, 1998.
- E. Gould. Neonatal vocalizations in bats of eight genera. *J. Mammal.*, 56:15–29, 1975.
- E. Gould, N. K. Woolf, and D. C. Turner. Double-note communication calls in bats: occurrence in three families. *J. Mammal.*, 54:998–1001, 1973.

BIBLIOGRAPHY

- S. D. Gray, I. R. Titze, F. Alipour, and T. H. Hammond. Biomechanical and histologic observations of vocal fold fibrous proteins. *Ann. Otol. Rhinol. Laryngol.*, 109:77–85, 2000.
- D. R. Griffin. *Listening in the Dark: The Acoustic Orientation of Bats and Men*. Yale University Press, New Haven, Connecticut, 1958.
- D. R. Griffin, F. A. Webster, and C. R. Michael. The echolocation of flying insects by bats. *Anim. Behav.*, 8:141–154, 1960.
- P. Griffiths. *Modern music: The Avant-Garde since 1945*. George Braziller, New York, 1981.
- J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields*. Springer-Verlag, 1983.
- T. Haji, K. Mori, K. Omori, and N. Isshiki. Experimental studies on the viscoelasticity of the vocal fold. *Acta Otolaryngol.*, 112:151–159, 1992a.
- T. Haji, K. Mori, K. Omori, and N. Isshiki. Mechanical properties of the vocal fold. stress-strain studies. *Acta Otolaryngol.*, 112:559–565, 1992b.
- B. Hammarberg. High-Speed Observations of Diplophonic Phonation. In O. Fujimura and M. Hirano, editors, *Vocal Fold Physiology*, chapter 21, pages 343–345. Singular Publisher Group, San Diego, Ca, 1995.
- G. H. Hardy and E. M. Wright. *Theory of Numbers*. Oxford University Press, Oxford, 1938.
- D. F. N. Harrison. *The Anatomy and Physiology of the Mammalian Larynx*. Cambridge University Press, New York, 1995.
- D. M. Hartl, S. Hans, J. Vaissiere, M. Riquet, and D. F. Brasnu. Objective voice quality analysis before and after onset of unilateral vocal fold paralysis. *J. Voice*, 15:351–361, 2001.
- D. J. Hartley and R. A. Suthers. The acoustics of the vocal tract in the horseshoe bat, *Rhinolophus hildebrandti*. *J. Acoust. Soc. Am.*, 84:1201–1213, 1988.
- D. J. Hartley, K. A. Campbell, and R. A. Suthers. The acoustic behavior of the fish-catching bat, *Noctilio leporinus*, during prey capture. *J. Acoust. Soc. Am.*, 86:8–27, 1989.
- M. Hast. The larynx of roaring and non-roaring cats. *Journal of Anatomy*, 163: 117–121, 1971.

BIBLIOGRAPHY

- U. Häusler. Vocalization-correlated respiratory movements in the squirrel monkey. *J. Acoust. Soc. Am.*, 108:1443–1450, 2000.
- H. Herzel. Bifurcations and chaos in voice signals. *Appl. Mech. Rev.*, 46:399–413, 1993.
- H. Herzel and C. Knudsen. Bifurcations in a vocal fold model. *Nonlinear Dynamics*, 7:53–64, 1995.
- H. Herzel and R. Reuter. Whistle register and biphonation in a child’s voice. *Folia Phoniatr. Logop.*, 49:216–224, 1997.
- H. Herzel and J. Wendler. Evidence of chaos in phonatory samples. In ESCA, editor, *Proc. EUROSPEECH (Genova)*, pages 263–266, 1991.
- H. Herzel, D. A. Berry, I. R. Titze, and M. Saleh. Analysis of vocal disorders with methods from nonlinear dynamics. *J. Speech Hear. Res.*, 37:1008–1019, 1994.
- H. Herzel, D. A. Berry, I. R. Titze, and I. Steinecke. Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling. *Chaos*, 5:30–34, 1995.
- H. Herzel, J. Holzfuss, Z. J. Kowalik, B. Pompe, and R. Reuter. Detecting bifurcations in voice signals. In H. Kantz, J. Kurths, and G. Mayer-Kress, editors, *Nonlinear Analysis of Physiological Data*, pages 325–344. Springer, Berlin, 1998.
- M. Hess, M. Gross, and H. Herzel. Hochgeschwindigkeitsaufnahmen von Schwingungsmoden der Stimmlippen. *Otorhinolaryngol. Nova*, 4:307–312, 1994.
- M. Hirano. Morphological structure of the vocal cord as a vibrator and its variations. *Folia phoniatr.*, 26:89–94, 1974.
- J. F. Holzrichter, G. C. Burnett, L. C. Ng, and W. A. Lea. Speech articulator measurements using low power em-wave sensors. *J. Acoust. Soc. Am.*, 103:622–625, 1998.
- A. Homler. *Five Voices*. Compact Disc, Intakt, Zürich, 1989.
- I. Honjo, S. Tanaka, and M. Tanabe. Pathogenesis of protruded false vocal fold. *Arch. Otolaryngol.*, 111:398–399, 1985.
- K. Ishizaka and J. L. Flanagan. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Syst. Techn. J.*, 51:1233–1268, 1972.
- K. Ishizaka and N. Isshiki. Computer simulation of pathological vocal-cord vibration. *J. Acoust. Soc. Am.*, 60:1193–1198, 1976.

BIBLIOGRAPHY

- N. Isshiki. *Functional Surgery of the Larynx*. Kyoto University, Kyoto, Japan, 1977.
- E. A. Jackson. *Perspectives of nonlinear dynamics*. Cambridge University Press, 1989.
- K. Jensen. Extensions of mind and voice. *Composer*, 2:13–17, 1979.
- J. J. Jiang and Y. Zhang. Nonlinear dynamic analysis of speech from pathological subjects. *Electronics Letters*, 38:294–295, 2002a.
- J. J. Jiang and Y. Zhang. Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds. *J. Acoust. Soc. Am.*, 112:2127–2133, 2002b.
- J. J. Jiang, I. R. Titze, D. B. Wexler, and S. D. Gray. Fundamental frequency and amplitude perturbation in reconstructed canine vocal folds. *Ann. Otol. Rhinol. Laryngol.*, 103:145–148, 1994.
- J. J. Jiang, C. E. Diaz, and D. G. Hanson. Finite element modeling of vocal fold vibration in normal phonation and hyperfunctional dysphonia: Implications for the pathogenesis of vocal nodules. *Ann. Otol. Rhino. Laryngol.*, 107:603–610, 1998.
- J. J. Jiang, C. I. Chang, J. R. Raviv, S. Gupta, F. M. Banzali, and D. G. Hanson. Quantitative study of mucosal wave via videokymography in canine larynges. *Laryngoscope*, 110:1567–1573, 2000.
- J. J. Jiang, A. G. Shah, M. M. Hess, K. Verdolini, F. M. Banzali, and D. G. Hanson. Vocal fold impact stress analysis. *J. Voice*, 15:4–14, 2001a.
- J. J. Jiang, Y. Zhang, and J. Stern. Modeling of chaotic vibrations in symmetric vocal folds. *J. Acoust. Soc. Am.*, 110:2120–2128, 2001b.
- G. Jones. Scaling of echolocation call parameters in bats. *J. Exp. Biol.*, 202:3359–3367, 1999.
- U. Jürgens. A study of the central control of vocalization using the squirrel monkey. *Medical Engineering and Physics*, 24:473–477, 2002.
- U. Jürgens. Reinforcing concomitants of electrically elicited vocalizations. *Exp. Brain Res.*, 26:203–214, 1976a.
- U. Jürgens. Projections from the cortical larynx area in the squirrel monkey. *Exp. Brain Res.*, 25:401–411, 1976b.

BIBLIOGRAPHY

- U. Jürgens. The squirrel monkey as an experimental model in the study of cerebral organization of emotional vocal utterances. *Eur. Arch. Psychiatry Neurol. Sci.*, 236:40–43, 1986.
- U. Jürgens. Central control of monkey calls. In D. Todt, P. Geodeking, and D. Symmes, editors, *Primate Vocal Communication*, pages 162–167. Springer-Verlag, Berlin, 1988.
- U. Jürgens. Neuronal control of mammalian vocalization, with special reference to the squirrel monkey. *Naturwissenschaften*, 85:376–388, 1998.
- U. Jürgens and P. Zwirner. Individual hemispheric asymmetry in vocal fold control of the squirrel monkey. *Behav. Brain Res.*, 109:213–217, 2000.
- U. Jürgens, M. Maurus, D. Ploog, and P. Winter. Vocalization in the squirrel monkey (*Saimiri sciureus*) elicited by brain stimulation. *Exp. Brain Res.*, 4: 114–117, 1967.
- U. Jürgens, M. Hast, and R. Pratt. Effects of laryngeal nerve transection on squirrel monkey calls. *J. Comp. Physiol. [A]*, 123:23–29, 1978.
- U. Jürgens, L. Ehrenreich, and N. C. D. Lanerolle. 2-deoxyglucose uptake during vocalization in the squirrel monkey brain. *Behavioural Brain Research*, 136: 605–610, 2002.
- T. Kaneko, H. Asano, J. Naito, N. Kobayashi, K. Hayashi, and T. Kitamura. Bio-mechanics of the vocal cords – on damping ratio. *J. Jpn. Soc. Bronchoesophagol.*, 25:133–138, 1972.
- T. Kaneko, K. Ushida, H. Suzuki, K. Komatsu, T. Kanesaka, N. Kobayashi, and J. Naito. Mechanical properties of the vocal fold: Measurement in vivo. In K. N. Stevens and M. Hirano, editors, *Vocal Fold Physiology*, pages 365–376. University of Tokyo Press, 1981.
- T. Kaneko, K. Komatsu, H. S. T. Kanesaka, T. Numata, and J. Naito. Mechanical properties of the human vocal fold - resonance characteristics in living humans and in excised larynges. In I. R. Titze and R. C. Scherer, editors, *Vocal Fold Physiology*, pages 305–317. The Denver Center of Performing Arts, 1983.
- T. Kaneko, T. Masuda, A. Shimada, H. Suzuki, K. Hayasaki, and K. Komatsu. Resonance characteristics of the human vocal fold in vivo and in vitro by impulse excitation. In T. Baer, C. Sasaki, and K. Harris, editors, *Vocal Fold Physiology*, pages 349–365. Little Brown Boston, 1986.

BIBLIOGRAPHY

- D. Kaplan and L. Glass. *Understanding Nonlinear Dynamics*. Springer, Berlin, 1995.
- W. Kaufman. *Tibetan buddhist chant*. Indiana University Press, Bloomington, 1975.
- D. Kavasch. An introduction to extended vocal techniques: some compositional aspects and performance problems. In *Reports from the center*, volume 1. La Jolla: Center for Music Experiment, University of California-San Diego, 1980.
- M. Kimura. "Subharmonics:" an extended technique for the violin. *J. Acoust. Soc. Am.*, 97:3270, 1995.
- S. Kiritani, H. Hirose, and H. Imagawa. High-speed digital image analysis of vocal fold vibration in diplophonia. *Speech Communication*, 13:23–32, 1993.
- S. Kiritani, H. Imagawa, and H. Hirose. Vocal fold vibrations associated with involuntary voice changes in certain pathological cases. In O. Fujimura and M. Hirano, editors, *Vocal Fold Physiology*, pages 269–281. Singular Publisher Group, San Diego, Ca, 1995.
- A. Kirzinger and U. Jürgens. The effects of brainstem lesions on vocalization in the squirrel monkey. *Brain Res.*, 358:150–162, 1985.
- D. H. Klatt and L. C. Klatt. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.*, 87:820–857, 1990.
- K. J. Kohler. Articulatory reduction in German spontaneous speech. In ESCA, editor, *Proc. 4th Speech Prod. Seminar (Autrans)*, pages 1–4. Singular Publishing Group, San Diego, 1996.
- M. Kotby, I. Titze, M. Saleh, and D. Berry. Fundamental frequency stability in functional dysphonia. *Acta Otolaryngol.*, 113:439–444, 1993.
- A. Kumar and S. K. Mullick. Nonlinear dynamical analysis of speech. *J. Acoust. Soc. Am.*, 100:615–629, 1996.
- T. Kusuyama, H. Fukuda, A. Shiotani, H. Nakagawa, and J. Kanzaki. Analysis of vocal fold vibration by x-ray stroboscopy with multiple markers. *Otolaryngol. Head Neck Surg.*, 124:317–322, 2001.
- R. Laje and G. B. Mindlin. Diversity within a birdsong. *Phys. Rev. Lett.*, 89: 288102, 2002.

BIBLIOGRAPHY

- R. Laje, T. J. Gardner, and G. B. Mindlin. Neuromuscular control of vocalizations in birdsong: A model. *Phys. Rev. E*, 65:051921, 2002.
- W. C. Lancaster and J. R. Speakman. Variations in respiratory muscle activity during echolocation when stationary in three species of bat (microchiroptera: Vespertilionidae). *J. Exp. Biol.*, 204:4185–4197, 2001.
- P. S. Landa and M. G. Rosenblum. Time series analysis for system identification and diagnostics. *Physica D*, 48:232–254, 1991.
- J. Large and T. Murry. Studies of extended vocal techniques: safety. *NATS Bulletin*, 34:30–33, 1978.
- J. Large and T. Murry. Observation on the nature of Tibetan chant. *J. Res. Singing*, 1979.
- W. Lauterborn. Acoustic turbulence. In D. Sette, editor, *Frontiers in Physical Acoustics*, pages 123–144. North-Holland, Amsterdam, 1986.
- W. Lauterborn and U. Parlitz. Methods of chaos physics and their application to acoustics. *J. Acoust. Soc. Am.*, 84:1975–1993, 1988.
- W. Lauterborn and E. Suchla. Bifurcation superstructure in a model of acoustic turbulence. *Phys. Rev. Lett.*, 53:2304–2307, 1984.
- M.-H. Lee, J.-N. Lee, and K.-S. Soh. Chaos in segments from Korean traditional singing and Western singing. *J. Acoust. Soc. Am.*, 103:1175–1182, 1998.
- T. Levin. *The hundred thousand fools of god: Musical travels in Central Asia*. Indiana University Press, Bloomington, 1996.
- J. Liliencrantz. Numerical simulation of glottal flow. In J. Gauffin and B. Hammarberg, editors, *Vocal Fold Physiology: Acoustics, perception and physiological aspects of voice mechanisms*, pages 99–104. Singular Publishing Group, San Diego, CA, 1991.
- H. Liska-Aurbacher. Die Stimme kann mehr als Singen und Sprechen. preprint, 2000.
- N. J. C. Lous, G. C. J. Hofmans, R. N. Veldhuis, and A. Hirschberg. A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design. *Acta Acustica*, 84:1135–1150, 1998.
- J. C. Lucero. Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region. *J. Acoust. Soc. Am.*, 94:3104–3111, 1993.

BIBLIOGRAPHY

- J. C. Lucero. A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset. *J. Acoust. Soc. Am.*, 105:423–431, 1999.
- J. C. Lucero and T. Gotoh. On the threshold pressure and the minimum sustaining pressure in the vocal fold oscillation. *J. Acoust. Soc. Jpn.*, 14:213–214, 1993.
- C. Maganza, R. Caussé, and F. Laloë. Bifurcations, period doublings and chaos in clarinetlike systems. *Europhys. Lett.*, 1:295–302, 1986.
- W. A. Marasovich, H. S. Gopal, S. E. Gerber, and W. S. Gibson. Diplophonia in a neonate. *Int. J. of Pediatric Otorhinolaryngology*, 25:227–234, 1993.
- T. Masuda. Mechanical properties of the human vocal fold in vivo—resonance characteristics by a single rectangular pulse excitation. *Nippon Jibiinkoka Gakkai Kaiho*, 89:763–773, 1986.
- L. Matassini, R. Hegger, H. Kantz, and C. Manfredi. Analysis of vocal disorders in a feature space. *Med. Eng. Phys.*, 22:413–418, 2000.
- K. Mathiak, U. Klose, H. Ackermann, I. Hertrich, W. E. Kincses, and W. Grodd. Stroboscopic articulography using fast magnetic resonance imaging. *Int. J. Lang. Commun. Disord.*, 35:419–425, 2000.
- M. Mazo. Laments made visible: a study of paramusical elements in Russian lament. In B. Yung and J. S. C. Lam, editors, *Themes and Variations*, pages 164–211. Harvard University/Chinese University of Hong Kong, Harvard/Hong Kong, 1994.
- M. Mazo, D. Ericson, and T. Harvey. Emotion and expression: temporal data on voice quality in Russian lament. In O. Fujimura and M. Hirano, editors, *Vocal Fold Physiology: Voice quality control*, pages 173–178. Singular Publishing Group, San Diego, 1995.
- J. C. McKinney. *The diagnosis and correction of vocal faults*. Broadman, Nashville, 1982.
- W. Mende, H. Herzel, and K. Wermke. Bifurcations and chaos in newborn cries. *Phys. Lett. A*, 145:418–424, 1990.
- P. Mergell. *Nonlinear dynamics of phonation - highspeed glottography and biomechanical modeling of vocal fold oscillations*. Phd Thesis, Shaker Verlag, Aachen, 1998.

BIBLIOGRAPHY

- P. Mergell and H. Herzel. Observation and modeling of pitch-formant resonances. In T. Wittenberg, P. Mergell, M. Tigges, and U. Eysholdt, editors, *Advances in Quantitative Laryngoscopy*, pages 119–125, Göttingen, 1997a. Verlag Abteilung Phoniatrie.
- P. Mergell and H. Herzel. Modelling biphonation – the role of the vocal tract. *Speech Communication*, 22:141–154, 1997b.
- P. Mergell, H. Herzel, T. Wittenberg, M. Tigges, and U. Eysholdt. Phonation onset: Vocal fold modeling and high-speed glottography. *J. Acoust. Soc. Am.*, 104:464–470, 1998.
- P. Mergell, T. W. Fitch, and H. Herzel. Modeling the role of non-human vocal membranes in phonation. *J. Acoust. Soc. Am.*, 105:2020–2028, 1999.
- P. Mergell, H. Herzel, and I. Titze. Irregular vocal fold vibration – High-speed observation and modeling. *J. Acoust. Soc. Am.*, 108:2996–3002, 2000.
- D. Michaelis, T. Gramss, and H. W. Strube. Glottal-to-noise excitation ratio - a measure for describing pathological voices. *Acustica/Acta acustica*, 83:700–706, 1997.
- J. Miller, J. Pereira, and D. Thomas. Fluid-flow through the larynx channel. *J. Sound Vib.*, 121:277–290, 1988.
- Y. Min, F. Alipour, and I. R. Titze. Stress-strain response of the human vocal ligament. *Annals of Otology, Rhinology, and Laryngology*, 104:563–569, 1995.
- F. Miranda. *Las voces de la voz*. Compact Disc, U. M. Unio Musics, Mallorca, 1992.
- L. Mongeau, N. Franchek, C. H. Coker, and R. A. Kubli. Characteristics of a pulsating jet through a small modulated orifice, with application to voice production. *J. Acoust. Soc. Am.*, 102:1121–1133, 1997.
- P. Moore, F. D. White, and H. von Leden. Ultra high speed photography in laryngeal physiology. *J. Speech Hear. Disord.*, 27:165–170, 1962.
- C. F. Moss, D. Redish, C. Gounden, and T. H. Kunz. Ontogeny of vocal signals in the little brown bat, *Myotis lucifugus*. *Anim. Behav.*, 54:131–141, 1997.
- D. Moss. *Five Voices*. Compact Disc, Intakt, Zürich, 1989.
- H. Muta and T. Baer. A pitch-synchronous analysis of hoarseness in running speech. *J. Acoust. Soc. Am.*, 84:1292–1301, 1988.

BIBLIOGRAPHY

- S. S. Narayanan and A. A. Alwan. A nonlinear dynamical systems analysis of fricative consonants. *J. Acoust. Soc. Am.*, 97:2511–2524, 1995.
- S. Nasri, J. A. Sercarz, and G. S. Berke. Noninvasive measurement of traveling wave velocity in the canine larynx. *Ann. Otol. Rhinol. Laryngol.*, 103:758–766, 1994.
- T. Nawka, L. Anders, and J. Wendler. Die auditive Beurteilung heiserer Stimmen nach dem RBH-System. *Sprache Stimme Gehör*, 18:130–133, 1994.
- V. E. Negus. *The Comparative Anatomy and Physiology of the Larynx*. Hafner Publishing Company, New York, 1949.
- J. Neubauer, P. Mergell, U. Eysholdt, and H. Herzel. Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes. *J. Acoust. Soc. Am.*, 110:3179–3192, 2001.
- J. Neubauer, M. Edgerton, and H. Herzel. Nonlinear phenomena in contemporary vocal music. *J. Voice*, 18:1–12, 2004.
- R. M. Newell. Writing for singers in the sixties. D.M.A. Thesis, University of California-San Diego, 1970.
- J. Newman. Squirrel monkey communication. In L. A. Rosenblum and C. L. Coe, editors, *Handbook of Squirrel Monkey Research*, pages 99–125. Plenum Press, New Jersey, 1985.
- J. D. Newman, H. J. Smith, and G. Talmage-Riggs. Structural variability in primate vocalizations and its functional significance: an analysis of squirrel monkey chuck calls. *Folia Primatol. (Basel)*, 40:114–124, 1983.
- N. Nonomura, S. Seki, M. Kawana, T. Okura, and Y. Nakano. Acquired airway obstruction caused by hypertrophic mucosa of the arytenoids and aryepiglottic folds. *Am. J. of Otolaryngology*, 17:71–74, 1996.
- G. R. North, T. L. Bell, and R. F. Cahalan. Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review*, 110:699–706, 1982.
- A. Novick and D. R. Griffin. Laryngeal mechanisms in bats for the production of orientation sounds. *J. Exp. Zool.*, 148:125–145, 1961.
- T. Numata. Mechanical properties of the vocal fold in excised human larynges—resonance characteristics by a single rectangular pulse excitation. *Nippon Jibi-inkoka Gakkai Kaiho*, 88:853–867, 1985.

BIBLIOGRAPHY

- M. Ouaknine, R. Garrel, and A. Giovanni. Separate detection of vocal fold vibration by optorelectometry: a study of biphonation on excised porcine larynges. *Folia Phoniatr. Logop.*, 55:28–38, 2003.
- X. Pelorson, A. Hirschberg, R. R. van Hassel, and A. P. J. Wijnands. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. *J. Acoust. Soc. Am.*, 96:3416–3431, 1994.
- A. L. Perlman, I. R. Titze, and D. S. Cooper. Elasticity of canine vocal fold tissue. *J Speech Hear Res*, 27:212–219, 1984.
- M. Petursson and J. Neppert. *Elementarbuch der Phonetik*. Buske, Hamburg, 1991.
- R. Rachele. *Overtone singing study guide*. Cryptic Voices Productions, Amsterdam, 1997.
- T. Riede, G. Böhme, R. Frey, T. Fitch, M. L. East, H. Hofer, and H. Herzel. Canids and hyaenas possess morphological structures that could be responsible for nonlinear phenomena during vocalization. *Advances in Ethology (Suppl.)*, 35:63, 2000a.
- T. Riede, H. Herzel, D. Mehwald, W. Seidner, E. Trumler, G. Bohme, and G. Tembrock. Nonlinear phenomena in the natural howling of a dog-wolf mix. *J. Acoust. Soc. Am.*, 108:1435–1442, 2000b.
- T. Riede, H. Herzel, K. Hammerschmidt, L. Brunnberg, and G. Tembrock. The harmonic-to-noise ratio applied to dog barks. *J. Acoust. Soc. Am.*, 110:2191–2197, 2001.
- C. N. Riviere, R. S. Rader, and N. V. Thakor. Adaptive canceling of physiological tremor for improved precision in microsurgery. *IEEE Trans. Biomed. Eng.*, 45:839–846, 1998.
- J. B. Robb and J. Saxman. Acoustic observations in young children’s vocalizations. *J. Acoust. Soc. Am.*, 83:1876–1882, 1988.
- R. Rübsamen and M. Betz. Control of echolocation pulses by neurons of the nucleus ambiguus in the rufous horseshoe bat, *Rhinolophus rouxi*. i. single unit recordings in the ventral motor nucleus of the laryngeal nerves in spontaneously vocalizing bats. *J. Comp. Physiol. A*, 159:675–687, 1986.

BIBLIOGRAPHY

- A. K. Saadah, N. P. Galatsanos, D. Bless, and C. A. Ramos. Deformation analysis of the vocal folds from videostroboscopic image sequences of the larynx. *J. Acoust. Soc. Am.*, 103:3627–3641, 1998.
- C. Sachs. *The wellsprings of music*. The Hague, M. Nijhoff, 1962.
- R. C. Scherer, I. R. Titze, and J. F. Curtis. Pressure-flow relationships in two models of the larynx having rectangular glottal shapes. *J. Acoust. Soc. Am.*, 73:668–676, 1983.
- R. C. Scherer, D. Shinwari, K. J. D. Witt, C. Zhang, B. R. Kucinski, and A. A. Afjeh. Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. *J. Acoust. Soc. Am.*, 109:1616–1630, 2001.
- D. Schnebel. Sprech- und Gesangsschule. *Melos*, 4:198–206, 1972.
- H. U. Schnitzler, E. Kalko, L. Miller, and A. Surlykke. The echolocation and hunting behavior of the bat, *Pipistrellus kuhli*. *J. Comp. Physiol. A*, 161:267–274, 1987.
- J. Schoentgen. Stochastic models of jitter. *J. Acoust. Soc. Am.*, 109:1631–1650, 2001.
- M. A. Schön-Ybarra. A comparative approach to the nonhuman primate vocal tract: Implications for sound production. In E. Zimmerman and J. D. Newman, editors, *Frontiers in Primate Vocal Communication*. Plenum Press, New York, 1995.
- E. Schönhärl. *Die Stroboskopie in der praktischen Laryngologie*. Thieme Verlag, Stuttgart, 1960.
- D. Schott. Quantitative analysis of the vocal repertoire of squirrel monkeys (*Saimiri sciureus*). *Z. Tierpsychol.*, 38:225–250, 1975.
- H. G. Schuster. *Deterministic chaos: an introduction*. VCH Verlagsgesellschaft, Weinheim, Basel, Cambridge, 1988.
- W. S. Selbie, S. L. Gewalt, and C. L. Ludlow. Developing an anatomical model of the human laryngeal cartilages from magnetic resonance imaging. *J. Acoust. Soc. Am.*, 112:1077–1090, 2002.
- D. Shinwari, R. C. Scherer, K. J. DeWitt, and A. A. Afjeh. Flow visualization and pressure distributions in a model of the glottis with a symmetric and oblique divergent angle of 10 degrees. *J. Acoust. Soc. Am.*, 113:487–497, 2003.

BIBLIOGRAPHY

- J. A. Simmons, K. M. Eastman, S. S. Horowitz, M. J. O'Farrell, and D. N. Lee. Versatility of biosonar in the big brown bat, *Eptesicus fuscus*. *Acoustics Research Letters Online*, 2:43–48, 2001.
- P. Sirviö and K. Michelsson. Sound-spectrographic cry analysis of normal and abnormal newborn infants. *Folia phoniatr.*, 28:161–173, 1976.
- S. H. Sloan, G. S. Berke, and B. R. Gerratt. Effect of asymmetric vocal fold stiffness on traveling wave velocity in the canine larynx. *Otolaryngol. Head Neck Surg.*, 107:516–526, 1992.
- S. H. Sloan, G. S. Berke, B. R. Gerratt, J. Kreiman, and M. Ye. Determination of vocal fold mucosal wave velocity in an in vivo canine model. *Laryngoscope*, 103: 947–953, 1993.
- H. Smith, K. Stevens, and R. Tomlinson. On an unusual mode of chanting by certain Tibetan lamas. *J. Acoust. Soc. Am.*, 41:1262–1264, 1967.
- H. J. Smith, J. D. Newman, H. J. Hoffman, and K. Fetterly. Statistical discrimination among vocalizations of individual squirrel monkeys (*Saimiri sciureus*). *Folia Primatol. (Basel)*, 37:267–279, 1982.
- M. E. Smith, G. S. Berke, B. R. Gerratt, and J. Kreiman. Laryngeal paralyses: Theoretical considerations and effects on laryngeal vibration. *J. Speech Hear. Res.*, 35:545–554, 1992.
- T. Smyth and J. O. Smith. The syrinx: Nature's hybrid wind instrument. *J. Acoust. Soc. Am.*, 112:2240, 2002.
- D. Starck and R. Schneider. Respirationsorgane. In H. Hofer, A. H. Schultz, and D. Starck, editors, *Primatologia III/2*, pages 165–187. S. Karger, Basel, 1960.
- I. Steinecke and H. Herzel. Bifurcations in an asymmetric vocal fold model. *J. Acoust. Soc. Am.*, 97:1571–1578, 1995.
- B. H. Story and I. R. Titze. Voice simulations with a body-cover model of the vocal folds. *J. Acoust. Soc. Am.*, 97:1249–1259, 1995.
- B. H. Story, I. R. Titze, and E. A. Hoffman. Vocal tract area functions for an adult female speaker based on volumetric imaging. *J. Acoust. Soc. Am.*, 104:471–487, 1998.
- B. H. Story, I. R. Titze, and E. A. Hoffman. The relationship of vocal tract shape to three voice qualities. *J. Acoust. Soc. Am.*, 109:1651–1650, 2001.

BIBLIOGRAPHY

- D. Stratos. *Cantare la voce*. Compact Disc, Cramps Records, Milano, 1978.
- S. H. Strogatz. *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Perseus Books, Cambridge MA, 1994.
- J. Sundberg. Vocal fold vibration patterns and modes of phonation. *Folia Phoniatr. Logop.*, 47:218–228, 1995.
- R. A. Suthers. The production of echolocation signals by bats and birds. In P. E. Nachtigall and P. W. B. Moore, editors, *Animal Sonar: Processes and Performance*, pages 23–45. Plenum Press, New York, 1988.
- R. A. Suthers and J. M. Fattu. Mechanisms of sound production in echolocating bats. *American Zoologist*, 13:1215–1226, 1973.
- R. A. Suthers and J. M. Fattu. Selective laryngeal neuroanatomy and the control of phonation by the echolocating bat, *Eptesicus*. *J. Comp. Physiol. A*, 145: 529–537, 1982.
- R. A. Suthers and D. Margoliash. Motor control of birdsong. *Current Opinion in Neurobiology*, 12:684–690, 2002.
- R. A. Suthers, S. P. Thomas, and B. J. Suthers. Respiration, wing-beat and ultrasonic pulse emission in an echolocating bat. *J. Exp. Biol.*, 56:37–48, 1972.
- R. A. Suthers, F. Goller, and R. S. Hartley. Motor stereotypy and diversity in songs of mimic thrushes. *J. Neurobiol.*, 30:231–245, 1996.
- L. Terrio and D. Schreiberweiss-Merin. Acoustic analysis of diplophonia: a follow-up report. *Perceptual and motor skills*, 77:914, 1993.
- S. L. Thibeault, S. D. Gray, D. M. Bless, R. W. Chan, and C. N. Ford. Histologic and rheologic characterization of vocal fold scarring. *J. Voice*, 16:96–104, 2002.
- G. Thoms and U. Jürgens. Role of the internal laryngeal nerve in phonation: an experimental study in the squirrel monkey. *Exp. Neurology*, 74:187–203, 1981.
- M. Tigges, T. Wittenberg, F. Rosanowski, and U. Eysholdt. High-speed imaging and image processing in voice disorders. In *Optical and Imaging Techniques for Biomonitoring II*, Wien, 1996. Proc. SPIE 2927.
- M. Tigges, P. Mergell, H. Herzel, T. Wittenberg, and U. Eysholdt. Observation and modelling glottal biphonation. *Acustica/Acta acustica*, 83:707–714, 1997.

BIBLIOGRAPHY

- M. Tigges, T. Wittenberg, P. Mergell, and U. Eysholdt. Imaging of vocal fold vibration by digital multi-plane kymography. *Computerized Medical Imaging and Graphics*, 23:323–330, 1999.
- I. R. Titze. The human vocal cords: A mathematical model I. *Phonetica*, 28: 129–170, 1973.
- I. R. Titze. The human vocal cords: A mathematical model II. *Phonetica*, 29: 1–21, 1974.
- I. R. Titze. On the mechanics of vocal-fold vibration. *J. Acoust. Soc. Am.*, 60: 1366–1380, 1976.
- I. R. Titze. Comments on the myoelastic - aerodynamic theory of phonation. *J. Speech Hear. Res.*, 23:495–510, 1980.
- I. R. Titze. The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am.*, 83:1536–1552, 1988.
- I. R. Titze. Mechanical stress in phonation. *NCVS Status and Progress Report*, 3: 291–301, 1993.
- I. R. Titze. *Principles of voice production*. Prentice-Hall, 1994a.
- I. R. Titze. Mechanical stress in phonation. *J. Voice*, 8:99–105, 1994b.
- I. R. Titze. Exploring the human voice with computer simulation. *NCVS Status and Progress Report*, 13:141–148, 1999.
- I. R. Titze and B. Story. Acoustic interactions of the voice source and the lower vocal tract. *J. Acoust. Soc. Am.*, 101:2234–2243, 1997.
- I. R. Titze and W. J. Strong. Normal modes in vocal cord tissues. *J. Acoust. Soc. Am.*, 57:736–744, 1975.
- I. R. Titze and D. T. Talkin. A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. *J. Acoust. Soc. Am.*, 66:60–74, 1979.
- I. R. Titze, R. Baken, and H. Herzel. Evidence of chaos in vocal fold vibration. In I. Titze, editor, *Vocal Fold Physiology: Frontiers in Basic Science*, pages 143–188, San Diego, 1993. Singular Publishing Group.
- I. R. Titze, N. P. Solomon, E. S. Luschei, and M. Hirano. Interference between normal vibrato and artificial stimulation of laryngeal muscles at near-vibrato rates. *J. Voice*, 8:215–223, 1994.

BIBLIOGRAPHY

- I. R. Titze, S. S. Schmidt, and M. R. Titze. Phonation threshold pressure in a physical model of the vocal fold mucosa. *J. Acoust. Soc. Am.*, 97:3080–3084, 1995.
- I. R. Titze, B. H. Story, G. C. Burnett, J. F. Holzrichter, L. C. Ng, and W. A. Lea. Comparison between electroglottography and electromagnetic glottography. *J. Acoust. Soc. Am.*, 107:581–588, 2000.
- I. Tokuda, T. Riede, J. Neubauer, M. J. Owren, and H. Herzel. Nonlinear analysis of irregular animal vocalizations. *J. Acoust. Soc. Am.*, 111:2908–2919, 2002.
- C. J. van As, M. Tigges, T. Wittenberg, B. M. R. O. de Coul, U. Eysholdt, and F. J. M. Hilgers. High-speed digital imaging of neoglottic vibration after total laryngectomy. *Arch. otolaryngol.*, 25:891–897, 1999.
- J. van den Berg. Myoelastic-aerodynamic theory of voice production. *J. Acoust. Soc. Am.*, 1:227–244, 1957.
- C. E. Vilain, X. Pelorson, A. Hirschberg, L. L. Marrec, W. O. Root, and J. Willems. Contribution to the physical modeling of the lips. Influence of the mechanical boundary conditions. *Acta Acustica*, 89:882–887, 2003.
- P. G. von Doersten, K. Izdebski, J. C. Ross, and R. M. Cruz. Ventricular dysphonia: a profile of 40 cases. *Laryngoscope*, 102:1296–1301, 1992.
- J. G. Švec and H. K. Schutte. Videokymography: High-Speed Line Scanning of Vocal Fold Vibration. *J. Voice*, 10:201–205, 1996.
- J. G. Švec, H. K. Schutte, and D. G. Miller. On pitch jumps between chest and falsetto registers in voice: data from living and excised larynges. *J. Acoust. Soc. Am.*, 106:1523–1531, 1999.
- J. G. Švec, J. Horáček, F. Šram, and J. Veselý. Resonance properties of the vocal folds: In vivo laryngoscopic investigation of the externally excited laryngeal vibrations. *J. Acoust. Soc. Am.*, 108:1397–1407, 2000.
- P. H. Ward, J. W. Sanders, R. Goldman, and G. P. Moore. Diplophonia. *Ann Otol. Rhino. Laryngol.*, 78:771–777, 1969.
- J. Wendler, W. Seidner, G. Kittel, and U. Eysholdt. *Lehrbuch der Phoniatrie und Pädaudiologie*. Georg Thieme Verlag, Stuttgart, New York, 1996.
- D. H. Whalen, B. Gick, M. Kumada, and K. Honda. Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic f₀. *J. Phonetics*, 27:125–142, 1999.

BIBLIOGRAPHY

- I. Wilden, H. Herzel, G. Peters, and G. Tembrock. Subharmonics, biphonation, and deterministic chaos in mammal vocalization. *Bioacoustics*, 9:171–196, 1998.
- P. Winter. The variability of peep and twit calls in captive squirrel monkeys (*Saimiri sciureus*). *Folia Primatol.*, 10:204–215, 1969.
- P. Winter. Observations on the vocal behaviour of free-ranging squirrel monkeys (*Saimiri sciureus*). *Z. Tierpsychol.*, 31:1–7, 1972.
- P. Winter, D. Ploog, and J. Latta. Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance. *Exp. Brain Res.*, 1:359–384, 1966.
- T. Wishart. *On sonic art*. Gordon and Breach, London, 1983.
- T. Wittenberg. Automatic motion extraction from laryngeal kymograms. In T. Wittenberg, P. Mergell, M. Tigges, and U. Eysholdt, editors, *Advances in Quantitative Laryngoscopy*, pages 21–28, Göttingen, 1997. Verlag Abteilung Phoniatrie.
- T. Wittenberg. *Wissensbasierte Bewegungsanalyse von Stimmlippenschwingungen anhand digitaler Hochgeschwindigkeitsaufnahmen*. Phd Thesis, Shaker Verlag, Aachen, 1998.
- T. Wittenberg, M. Moser, M. Tigges, and U. Eysholdt. Recording, processing and analysis of digital highspeed sequences in glottography. *Machine Vision and Applications*, 8:399–404, 1995.
- T. Wittenberg, P. Mergell, M. Tigges, and U. Eysholdt. Quantitative characterization of functional voice disorders using analysis of highspeed video and modeling. In *Proc's ICASSP-97, Apr 21-24, Munich, Germany*, volume 3, 1997.
- T. Wittenberg, T. Tigges, P. Mergell, and U. Eysholdt. Functional imaging of vocal fold vibration: Digital multislice high-speed kymography. *J. Voice*, 14: 422–442, 2000.
- J. G. Wong and D. A. Waters. The synchronisation of signal emission with wing-beat during the approach phase in soprano pipistrelles (*Pipistrellus pygmaeus*). *J. Exp. Biol.*, 204:575–583, 2001.
- G. E. Woodson. Configuration of the glottis in laryngeal paralysis. I: Clinical study. *Laryngoscope*, 103:1227–1234, 1993.
- E. Yanagi and T. V. McCaffrey. Study of vibratory pattern of the vocal folds in the excised canine larynx. *Arch Otolaryngol. Head Neck Surg*, 118:30–36, 1992.

BIBLIOGRAPHY

- E. Yumoto and Y. Kadota. Pliability of the vocal fold mucosa in relation to the mucosal upheaval during phonation. *Arch Otolaryngol Head Neck Surg*, 124: 897–902, 1998.
- E. Yumoto, W. J. Gould, and T. Baer. Harmonic-to-noise-ratio as an index of the degree of hoarseness. *J. Acoust. Soc. Am.*, 71:1544–1550, 1982.
- C. Zhang, W. Zhao, S. H. Frankel, and L. Mongeau. Computational aeroacoustics of phonation, Part II: Effects of flow parameters and ventricular folds. *J. Acoust. Soc. Am.*, 112:2147–2154, 2002a.
- Z. Zhang, L. Mongeau, and S. H. Frankel. Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes. *J. Acoust. Soc. Am.*, 112:1652–1663, 2002b.
- W. Zhao, C. Zhang, S. H. Frankel, and L. Mongeau. Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms. *J. Acoust. Soc. Am.*, 112:2134–2146, 2002.

Appendix A

Empirical Orthogonal Function Analysis

The method of empirical orthogonal (eigen)functions is used to extract independent spatio-temporal structures from complex spatio-temporal vibration patterns. It uses correlations between the time series of the vocal fold edge points to decompose the complex spatio-temporal vibration pattern into principal modes of vibration (Berry et al., 1994). In the literature, the terms “coherent structures”, “spatio-temporal structures”, “principal modes” or just “modes” have been used interchangeably (Aubry et al., 1991; Berry et al., 1994). The EOF method is also called principal components analysis, biorthogonal decomposition, Karhunen-Loève expansion, proper orthogonal decomposition, principal factor analysis, singular value analysis, and the singular spectrum analysis (see Berry et al., 1994). In principle, the idea of empirical orthogonal function analysis is to find an orthogonal coordinate system in state space for which all data points are decorrelated.

Here, this study is interested in independent spatio-temporal structures about the rest position of the vocal folds. Therefore, the excursion $d_k^{(\alpha)}(t_i)$ from the glottal midline is divided into a mean and a oscillatory component:

$$\delta_k^{(\alpha)}(t_i) = d_k^{(\alpha)}(t_i) - \bar{d}_k^{(\alpha)} \quad (\text{A.1})$$

Here, $k = 1, \dots, N$ indicates the kymogram number, $i = 1, \dots, M$ is the time index, and $\alpha \in \{left, right\}$ specifies the side of the vocal folds. The mean $\bar{d}_k^{(\alpha)} = \left\langle d_k^{(\alpha)}(t_i) \right\rangle_{t_i}$ represents the dynamic equilibrium of each vocal fold, where $\left\langle \cdot \right\rangle_{t_i}$ denotes a time average. The covariance matrix $\mathbf{C}^{(\alpha)}$ measures linear interdependencies of the fluctuations $\delta_k^{(\alpha)}(t_i)$ about the dynamic equilibrium at different

A. Empirical Orthogonal Function Analysis

locations k, l of each vocal fold:

$$C_{kl}^{(\alpha)} = \frac{1}{M} \sum_{i=1}^M \delta_k^{(\alpha)}(t_i) \delta_l^{(\alpha)}(t_i) = \left\langle \delta_k^{(\alpha)}(t_i) \delta_l^{(\alpha)}(t_i) \right\rangle_{t_i} \quad (\text{A.2})$$

with $k, l = 1, \dots, N$. The covariance matrix is a real, symmetric and positive semidefinite matrix. This can be diagonalized and has real and nonnegative eigenvalues:

$$\mathbf{C}^{(\alpha)} \Phi_l^{(\alpha)} = \lambda_l^{2(\alpha)} \Phi_l^{(\alpha)} \quad (\text{A.3})$$

The normalized eigenvectors $\Phi_l^{(\alpha)}$, $l = 1, \dots, N$, correspond to the empirical orthogonal functions (EOF). Sometimes they are called “topos” to indicate their topological meaning (Aubry et al., 1991). The corresponding eigenvalues $\lambda_l^{2(\alpha)}$ carry the weights of the normalized EOFs. They express the variance of the associated EOF about the dynamic equilibrium. The eigenfunctions define a set of orthogonal directions in the N -dimensional state space $\{\delta_k^{(\alpha)}\}$, and the eigenvalues measure the variance of the observed set of states projected onto their corresponding EOF-axis. The first eigenfunction is the direction in state space for which the variance is maximal.

The total variance E_{tot} of the centered set of states $\{\delta_k^{(\alpha)}(t_i)\}$ is given by the sum of the eigenvalues $\lambda_l^{2(\alpha)}$ of the EOFs:

$$E_{tot}^{(\alpha)} = \sum_{l=1}^N \lambda_l^{2(\alpha)} = \left\langle \delta_k^{(\alpha)}(t_i) \delta_k^{(\alpha)}(t_i) \right\rangle_{k, t_i} \quad (\text{A.4})$$

$\lambda_l^{2(\alpha)}$ is the variance of one single spatio-temporal structure about the dynamic equilibrium of the vocal fold α , with $\alpha \in \{\text{left}, \text{right}\}$. As the EOFs $\Phi_l^{(\alpha)}$ establish an orthonormal system, the oscillatory components $\delta_k^{(\alpha)}(t_i)$ can be written as a linear superposition:

$$\delta_k^{(\alpha)}(t_i) = \sum_{l=1}^N a_l^{(\alpha)}(t_i) \Phi_l^{(\alpha)}(k) \quad (\text{A.5})$$

The temporal expansion coefficients $a_l^{(\alpha)}(t_i)$ are determined by the projection of the oscillatory components onto the eigenfunctions:

$$a_l^{(\alpha)}(t_i) = \sum_{k=1}^N \delta_k^{(\alpha)}(t_i) \Phi_l^{(\alpha)}(k) \quad (\text{A.6})$$

Sometimes, the temporal coefficients are called “chronos”, as they express the chronological importance of the associated “topos” (Aubry et al., 1991). They

A. Empirical Orthogonal Function Analysis

also can be regarded as temporal eigenfunctions corresponding to their associated EOFs that can be thought of as spatial eigenfunctions. Finally, the oscillations of the vocal fold edges about their dynamical equilibrium can be decomposed as:

$$\delta_k^{(\alpha)}(t_i) = \sum_{l=1}^N \lambda_l^{(\alpha)} \Psi_l^{(\alpha)}(t_i) \Phi_l^{(\alpha)}(k) \quad (\text{A.7})$$

with $\Psi_l^{(\alpha)}(t_i)$ defined by

$$a_l^{(\alpha)}(t_i) = \lambda_l^{(\alpha)} \Psi_l^{(\alpha)}(t_i) \quad (\text{A.8})$$

To characterize different kinds of complex spatio-temporal patterns for different laryngeal pathologies, measures are proposed that provide overall information about the distribution of spatio-temporal vibration modes.

The eigenvalues $\lambda_l^{2(\alpha)}$ contain information about the distribution of variance in the state space $\{\delta_k^{(\alpha)}\}$ in the direction of the modes. The information entropy (Shannon's entropy) measures the degree of disorder of irregular spatio-temporal glottal contour patterns. It quantifies the shape of the eigenvalue distribution. The relative weights of each EOF,

$$p_l^{(\alpha)} = \lambda_l^{2(\alpha)} / E_{tot}^{(\alpha)} \quad (\text{A.9})$$

can be regarded as a probability distribution for the projection of an observation onto the directions of the EOFs $\Phi_l^{(\alpha)}$ in state space $\{\delta_k^{(\alpha)}\}$. Thus, the information entropy $S_{tot}^{(\alpha)}$ is an overall quantity for the degree of disorder of this distribution:

$$S_{tot}^{(\alpha)} = - \sum_{l=1}^N p_l^{(\alpha)} \log p_l^{(\alpha)} \quad (\text{A.10})$$

If only a single eigenvalue is nonzero, the information entropy is zero. Then, a single empirical function is present. Furthermore, $S_{tot}^{(\alpha)}$ will be small for a sharply peaked distribution $p_l^{(\alpha)}$; whereas for a broad distribution, $S_{tot}^{(\alpha)}$ will become large. In the limit of equidistributed EOFs, where all eigenvalues become equal, the information entropy reaches its global maximum $S_{max} = \log N$.

Appendix B

Experimental Observations on Vocal Membranes

B.1 Selected experimental studies in bats

In the literature there are numerous studies on various topics related to ultrasound phonation in bats. Many publications focus on the descriptive statistics of ultrasonic pulse emission. In particular they contain detailed data on center pulse frequencies, minimum and maximum frequency in frequency sweeps, and pulse rates during prey capture. Different bat species, including insect catching bats and plant- and nectar-feeding bats, have been used in these studies.

The topics in literature on bats related to this study include:

- Laryngeal mechanisms of echolocation
- Perceptual properties of hearing in bats
- Synchronization of respiration–wing beat–echolocation
- Radiotelemetry of prey catching bats
- Vocal tract acoustics
- Doppler compensation during prey capture
- Ontogeny of echolocation calls
- Prey capture success correlated with echolocation structure
- Echolocation signal design
- Individuality recognition

B. Experimental Observations on Vocal Membranes

In the following, a list of significant studies is given:

- A computational sensorimotor model of bat echolocation (Erwin et al., 2001)
- Echolocation behavior and prey-capture success in foraging bats: laboratory and field experiments on *Myotis daubentonii* (Britton and Jones, 1999)
- Regular pulse emission in some megachiropteran bats (Funakoshi et al., 1995)
- Double-note communication calls in bats: occurrence in three families (Gould et al., 1973)
- Neonatal vocalizations in bats of eight genera (Gould, 1975)
- The acoustics of the vocal tract in the horseshoe bat, *Rhinolophus hildebrandti* (Hartley and Suthers, 1988)
- Scaling of echolocation call parameters (Jones, 1999)
- Variations in respiratory muscle activity during echolocation when stationary in three species of bat (Microchiroptera: Vespertilionidae) (Lancaster and Speakman, 2001)
- Ontogeny of vocal signals in the little brown bat, *Myotis lucifugus* (Moss et al., 1997)
- Laryngeal mechanisms in bats for the production of orientation sounds (Novick and Griffin, 1961)
- Control of echolocation pulses by neurons of the nucleus ambiguus in the rufous horseshoe bat, *Rhinolophus rouxi*. I. Single unit recordings in the ventral motor nucleus of the laryngeal nerves in spontaneously vocalizing bats (Rübsamen and Betz, 1986)
- Laryngeal mechanisms in bats for the production of orientation sounds (Novick and Griffin, 1961)
- Versatility of biosonar in the big brown bat, *Eptesicus fuscus* (Simmons et al., 2001)
- Respiration, wing-beat and ultrasonic pulse emission in an echolocating bat (Suthers et al., 1972)
- Mechanisms of sound production in echolocating bats (Suthers and Fattu, 1973)

B. Experimental Observations on Vocal Membranes

- Selective laryngeal neuroanatomy and the control of phonation by the echolocating bat, *Eptesicus* (Suthers and Fattu, 1982)
- The synchronisation of signal emission with wingbeat during the approach phase in soprano pipistrelles (*Pipistrellus pygmaeus*) (Wong and Waters, 2001)

Selected papers relevant for biomechanical modeling are reviewed in some detail.

B.1.1 Subglottic pressure and the control of phonation by the echolocating bat, *Eptesicus*

Review of Fattu and Suthers (1981): During vocalizations of *Eptesicus*, subglottal pressure has been measured with an inserted tracheal cannula with pressure transducer. For FM pulses (frequency modulated pulses) with fundamental frequencies between 20 and 50 kHz, pulse durations of about 1 to 5 ms have been observed. Pulse repetition rates reached up to 200 Hz. The maximal sound pressure level was about 120 dB in front of the mouth. For single FM pulses, subglottal pressure was approx. 30 cmH₂O, and, for groups of pulses, subglottal pressure was found to be about 40 cmH₂O. During downward sweeping FM pulses, subglottal pressure decreased whereas during constant frequency (CF) portions pressure remained constant. For ascending FM portions the pressure either increased or stayed constant.

The authors discussed the role of the vocal folds and vocal membranes based on these measurements. They argued that vibratory and resistive structures have to be present in the bat larynx. The thin vocal membranes are thought to be unlikely to serve as glottal resistance for high subglottal pressures. They proposed a mechanism for pulse generation: After vocal fold adduction and closing of the glottis, subglottal pressure builds up during expiration. When a certain threshold is reached, the vocal folds slightly open until due to elastic recoil and aerodynamic driving forces the vocal membranes exhibit self-sustained oscillations for some hundred cycles. The vibration frequency is then determined by the tension that is exerted on the vocal membranes by the cricothyroid muscles. This hypothesis was further substantiated by denervation experiments of the cricothyroid muscle where the typical pulse frequencies were reduced into the audible range and the frequency modulation was eliminated (Griffin, 1958; Novick and Griffin, 1961; Suthers and Fattu, 1982). Furthermore, measurements of electrical activity in the superior laryngeal nerves and the major ventrolateral portion of the cricothyroid muscle showed muscle tension to be maximal just prior to pulse production and to decrease during phonation due to relaxation (Suthers and Fattu, 1973).

B.1.2 Mechanisms of sound production in echolocating bats

Review of Suthers and Fattu (1973): This study tried to reveal the mechanism of echolocative pulses at high repetition rates by bats. Vocalizations of the North American Vespertilionid bat *Eptesicus fuscus* showed pulse durations less than 1 to 10 ms. During FM downward sweeps, fundamental frequencies can decrease by about one octave, i.e. frequency halving, from 60 – 80 kHz to 30 – 40 kHz. Pulse rates can increase up to 200 Hz. Anatomically, these bats have unusually large cricoid and thyroid cartilages where the cricothyroid muscle is greatly hypertrophied.

Therefore, as a sound production mechanism, the authors proposed that the thin vocal membranes could be stretched by contraction of the cricothyroid muscles. The membranes are about 6 – 8 microns thick, and about 0.5 mm wide and 2 mm long. The vocal folds were thought of as not only sound generators, but also as airway resistance. They provide the glottal stop and thus control subglottic pressure during phonation.

The authors measured subglottal pressures in restrained awake bats that were electrically stimulated to vocalize. For single pulses, subglottal pressures could be determined as 25 to 40 cmH₂O before pulses, then dropping during pulses by about 20 cmH₂O within 5 ms. At high pulse repetition rates, several pulses were grouped within single respiratory cycles. Moreover, the peak subglottal pressure reached values up to 60 to 70 cmH₂O dropping by 20 – 45 cmH₂O during 2 – 3 ms of each pulse of grouped pulses.

The authors noted that, during stimulated vocalizations, the bats emitted sounds termed “protest cries” with substantial portions of wide-band frequency spectra and with large portions in the audible range. They also observed cries with transitions to irregular phonation where the cry begins with discrete FM components and then abruptly changed to “protest cries” of wide-band frequency spectrum.

Seeking insight into the innervation and control mechanisms of echolocative pulse emission, the authors conducted denervation experiments of the recurrent laryngeal nerve. This nerve innervates all intrinsic laryngeal muscles except the cricothyroid muscle. Denervation of the recurrent nerve showed little effect on ultrasonic pulses. After denervation of the cricothyroid muscles (accomplished by cutting the superior laryngeal nerves), frequencies of echolocative pulses were lowered to audible frequencies. FM pulses changed into pulses with relatively constant frequency. Interestingly, the pulse duration was relatively unaffected. The authors concluded that the cricothyroid muscle mainly exerts tension on the vocal membranes.

Using force transducer triodes, contraction of the cricothyroid muscle could be measured. It was found that the cricothyroid muscle of bats is well adapted for

B. Experimental Observations on Vocal Membranes

rapid contractions for pulse repetition rates up to 200 Hz. During single electrical motor nerve stimulation, contraction times of about 6.5 ms could be observed. The total contraction/relaxation time of the bat cricothyroid was about 12 – 16 ms. This is significantly longer than the interpulse interval at a maximal pulse repetition rate with 4 – 5 ms interpulse intervals. Using repetitive electrical stimulations of the motor nerve (stimulation rate of about 55 – 65 Hz), partial tetanus of the cricothyroid muscle was observed. Partial relaxation occurred between each stimulus. Partial relaxation ceased when the fusion frequency of 220 – 240 Hz was reached. The authors concluded that the partial muscle relaxation of the cricothyroid could be responsible for frequency modulations in FM pulse emitting bats.

B.1.3 The production of echolocation signals by bats and birds

Review of Suthers (1988): In this review paper on echolocation in frequency modulating (FM) and constant frequency (CF) bats, the mechanisms for pulsed echolocation calls were described. Important details for the presented vocal membrane model are summarized below.

Vocal membranes in bats consist of connective tissue. They are only several microns thick. Traverse cuts in vocal membranes left bats aphonic. Only faint clicks and sometimes fricative hissing sounds could be produced after this procedure.

Echolocating bats produce sonar pulses during the expiratory phase of the respiratory cycle. They use two different strategies for pulses. One pattern is characterized by low pulse repetition rates and long lasting pulses within one repetition (large duty cycle). The second pattern consists of short pulses and high pulse repetition rates.

During sonar pulses, subglottal pressure reaches values of about 20-50 cmH₂O (human conversation: about 8 cmH₂O and shouting: 20-30 cmH₂O). Vocal efficiency of about 7-16 percent (the ratio of acoustical and mechanical vibration energy to air flow energy) is about two orders of magnitude higher than in humans. The maximal subglottal pressure is thought to be limited by the blood pressure in the lung capillaries (about 43 cmH₂O) which is supposed to be higher during flight (increased heart rate).

The mechanical tension in vocal membranes is controlled by the cricothyroid muscle (CTM), which is hypertrophied in bats of the genus *Microchiroptera*. The CTM is innervated by the superior laryngeal nerve. All other intrinsic laryngeal muscles are innervated by the inferior (recurrent) laryngeal nerve. During sonar pulse trains electrical activity in the CTM is high before each vocalization. The downward FM sweep in FM bats is due to CTM relaxation and decreasing vocal membrane tension.

B. Experimental Observations on Vocal Membranes

After transection of the inferior laryngeal nerve, FM bats were still able to produce FM pulses. The maximal subglottal pressure was reduced which resulted in a reduced maximal sound pressure level. Due to the induced lack of timing of the glottal gate, abnormal rising FM pulses were observed. Instead of pulse trains, long-lasting vocalizations with sinusoidally varying fundamental frequency were found. The ventral thyroarytenoid muscle (ventral part of the vocalis muscle) shows high electrical activity synchronous to the CTM before phonation. It is innervated by the inferior laryngeal nerve. In FM bats it is thought to be a glottal adductor. It reduces the tension on vocal folds and its contraction bulges the vocal folds into the glottal space.

Transection of the superior laryngeal nerve in FM bats led to abnormal pulses. Pulse duration became highly variable, and frequency modulation vanished. The fundamental frequency decreased to about 8 kHz. Series of harmonics appeared.

These transection experiments show that in bats the CTM is used to control glottal resistance and vocal membrane tension. It adducts the vocal folds and tenses the vocal membranes. Thus, vocal membrane tension and glottal resistance are increased simultaneously before sonar pulses. During the FM downward sweep the CTM relaxes and simultaneously opens the glottis and decreases vocal membrane tension.

In bats with CF sonar pulses, denervation of the CTM resulted in a decrease of fundamental frequency from, e.g., 83 kHz to 12-42 kHz. The fundamental frequency became highly variable from pulse to pulse. Doppler shift typical for CF bats vanished after CTM denervation.

In bats using both CF and FM echolocation, the glottal resistance is held constant while expiratory muscles increase or decrease laryngeal airflow.

In summary, the sonar pulse type in bats (rising FM, descending FM, CF, or combinations) depends on the phase relationship between phonation onset/offset and the contraction/relaxation cycle of the cricothyroid muscle. Phonation can be terminated in two ways. Due to adduction of the vocal folds against the rising subglottal pressure, the glottis can be closed to stop pulse trains at high pulse repetition rates. Alternatively, the vocal membranes can abduct out of the phonatory position. This results in an increase of subglottal airflow and a decrease of subglottal driving pressure.

B.2 Selected experimental studies in primates

Squirrel monkeys have become the most extensively investigated species in studies on the central nervous control of vocalization because: they have a rich vocal repertoire, they are members of the primate order, their brain resembles the human brain anatomy to a much higher degree than that of the cat or rat, and they are

easy to handle for experiments (Jürgens, 2002). Moreover, squirrel monkey vocalization can be considered as a suitable model for the study of the central control of human nonverbal emotional vocal utterances such as laughing, moaning, crying, and jubilating (Jürgens, 1986; Jürgens et al., 2002). Different human nonverbal emotional vocal utterances (e.g., laughing, shrieking, moaning) and emotional intonation patterns (e.g., scolding, lamenting, caressing) can be shown to have acoustic and emotional counterparts in the vocal repertoire of the squirrel monkey.

Important byproducts of these studies are the vocalization repertoire, the corresponding behavioral context, and the neural control mechanisms. They are well documented and typically combined with spectrographic displays of calls (e.g., Boinski and Mitchell (1995, 1997); Newman et al. (1983); Schott (1975); Winter et al. (1966); Winter (1969, 1972)).

In the following, a few prominent articles on the role of vocal folds and vocal membranes for the call repertoire of primates will be reviewed.

B.2.1 Effects of laryngeal nerve transection on squirrel monkey calls

Review of Jürgens et al. (1978): This study bridges experimental work on neurophysiological innervation of phonation in bats and humans using the squirrel monkey as an animal “between” bats and humans with respect to the call repertoire. The authors interrupted systematically the two major innervation paths that control the intrinsic laryngeal muscles (e.g., cricothyroid, thyroarytenoid muscles) to discern the dynamic relevance of intrinsic muscles. When cutting the recurrens nerve, which innervates all intrinsic muscles apart from the cricothyroid muscle, low frequency containing calls were effected more structurally than high frequency calls. Rhythmical pulse sequences/click series at low pulse rates became arrhythmical while for broadband noise-like calls, the width of the main energy band decreased, and low frequency harmonic calls were superimposed with noise-like components. Wideband noise-like calls could even change to contain harmonic components, sometimes with a quasi-fundamental frequency of about 2 – 3 kHz. After superior laryngeal nerve transection (which innervates the cricothyroid muscle only) all calls with low frequency components remained unaffected. One interesting finding (from the viewpoint of the vocal membrane model discussed in this thesis) is the appearance of faint harmonic-like structures superimposed on the continuous frequency band typical for normal shrieking that showed up only during soft shrieking calls. High pitched constant-frequency calls decreased in fundamental frequency significantly and high pitched frequency-modulated calls disappeared.

Several conclusions can be drawn from this study. It can be assumed that the cricothyroid muscle plays major role in stiffening the vocal membranes. From

B. Experimental Observations on Vocal Membranes

human physiology it is known that the cricothyroid muscle elongates the vocal folds and narrows the glottis.

The production of rhythmical click sequences could be due to a dynamic effect rather than an individual neural muscle activation for each pulse. Therefore click series did not disappear completely after unilateral recurrens transection, but only became arrhythmic.

After bilateral recurrens transection, the click series were observed to decrease even more in sound intensity. Single clicks changed from wide-band, noise-like sounds to harmonic calls. After recurrens transection, a large prephonatory glottis and missing vocal fold contact could cause the vocal folds and membranes to be weakly driven by the airstream. This could lead to a more periodic behavior of the vibrating modes with small oscillation amplitudes reducing nonlinearity in the system.

Narrowing of frequency bands in broadband calls and the appearance of harmonic-like structures with quasi-fundamentals of several kHz could be an effect of the decrease of vocal fold contact due to an increase in glottal area. Then the nonlinear interaction decreases and the vocal membranes show a more periodic behavior.

Harmonic-like structures in broadband calls could be generated by an irregularly oscillating part of the vocal folds exciting the passively oscillating flaccid vocal membranes to vibrate in a periodic way.

B.2.2 Modes of vocal variation in Sykes's monkey (*Cercopithecus albogularis*) squeals

Review of Brown and Cannito (1995): This study uses laryngographic waveform measurements (electroglottogram EGG) to reveal mechanisms of sound generation in Sykes monkeys (*Cercopithecus albogularis*), an Old World primate of about 40–70 cm body length and 4–12 kg weight. Based on acoustical parameters cluster analysis differentiated four categories for Sykes's monkey squeals. Squeal calls are associated with high arousal levels, but not particularly with loud calls.

In general, EGG measurements showed that the range of variation in the surface area of the vocal folds making contact during phonation was approx. six times larger than in humans. The authors argued that the vocal membranes, the membranous extension of the vocal folds found in most nonhuman primates, could be the major reason for this drastic difference. Functionally, the valvelike sealing of the airway with vocal membranes generates high subglottal air pressure that could facilitate certain vocal gestures and stabilize the rib cage for upper body maneuvers in trees.

One class of squeal calls consisted of calls with broadband noise, low fundamental frequency (range approx. 600 Hz – 1400 Hz), and large amplitude oscillations of

B. Experimental Observations on Vocal Membranes

the EGG signal. A second class consisted of calls with small broadband noise, high fundamental frequency (range approx. 1100 Hz – 4800 Hz), and small amplitude oscillations of the EGG signal. As no significant evidences for irregular behavior of the vocal source could be found, broadband vocal elements could be explained by turbulence in the vocal tract.

These two classes were interpreted as being similar to modal (chest) and falsetto (loft) register in human phonation. The authors suggested that during the monkey's equivalent of the modal register, both vocal folds and vocal membranes make contact resulting in a large EGG signal. During the monkey's equivalent of the falsetto register only the smaller vocal membranes are thought to be in contact producing small EGG signals.

A third class of squeals was comprised of calls with multiple fundamental frequencies. The EGG signal contained superpositions of both high and low amplitude oscillations with different frequencies. Transitions between superpositions of two high frequency oscillations were shown where the oscillations with a frequency ratio of about 3 : 2 changed to oscillations with a high frequency component modulated by a strong low frequency oscillation (frequency ratio approx. 5). The EGG and sound spectrograms also support the hypothesis that the two frequency components have different time courses and thus could be classified as biphonation. The authors suggested that different regions of the larynx could vibrate simultaneously with certain frequency relations ("polyphonic source"). They hypothesize that vocal membranes and the principal portion of the vocal folds oscillate in different synchronization modes. Vocal membranes were found to vibrate in a 2 : 1 ratio related to the frequency of the vocal folds. In other cases the oscillation frequency of vocal membranes was much higher than that of the principal component of the vocal folds. The authors noted that no beating could be observed in this particular calls. This finding indicates that the frequency of the vocal membranes never came close to the frequency of the principal component of the vocal folds.

In the fourth class, abrupt transitions between narrow- and broad-band call components together with abrupt transitions between small and large EGG signals were observed. Additionally, the fundamental frequency abruptly changed between the lower and the higher frequency range. Irregular call sequences were interspersed with harmonically structured segments. This behaviors can be classified as register jumps and deterministic chaos with harmonic windows.

B.2.3 Voicing biomechanics and squirrel monkey (*Saimiri boliviensis*) vocal communication

Review of Brown et al. (2003): Vocal communication patterns in squirrel monkeys have been studied using excised larynx preparations to measure airflow and

B. Experimental Observations on Vocal Membranes

acoustical phonatory patterns *in vitro*. The vocal folds of squirrel monkeys are about 2 - 3 mm long. The larynx, trachea and associated tissues are about 8 times smaller than similar structures in humans. This makes excised larynx experiments technically quite difficult and prone to errors.

For the studied excised larynges, phonation onset pressure was approx. 5 - 8 cmH₂O. Because a subglottal pressure of about 40 cmH₂O irreversibly damaged one larynx preparation, an upper bound of a physiologically sensible pressure range was at this limit. At phonation onset hysteresis was observed. This behavior can be described in the framework of nonlinear dynamics as a subcritical Hopf bifurcation together with a saddle-node-bifurcation. After vibration onset, the subglottal pressure could be reduced until vibrations ceased at a lower pressure value than at phonation onset.

In vitro vibrational patterns could be grouped into a “periodic”, “biphonic”, “asymmetric” and “chaotic” regime. The “periodic regime” was characterized by a prominent fundamental frequency and a series of harmonic overtones. Within the vocalization repertoire of squirrel monkeys, “cackles” and other harmonically structured calls were considered related to the periodic regime of excised larynges. A shallow spectral slope associated with a small ratio of glottal opening time to glottal period (open quotient) was observed in this regime. Therefore the authors suggested that call audibility *in vivo* could be increased; and, given a fixed glottal flow rate, the number of glottal cycles per exhalation (utterance duration) could be increased.

The “biphonic regime” contained patterns with nonharmonically related peak frequencies. A variety of calls from the squirrel monkey’s vocalization repertoire could be associated with the biphonic regime, e.g., calls in the “twitter”, “trill”, “chuck”, and “kecker” classes. These show nonharmonic spectral bands and independent frequency contours of frequency bands. The authors hypothesized that in excised larynges, biphonation could be induced either by left-right asymmetries between vocal folds, desynchronization of vertical and horizontal vibratory modes, or desynchronization of vocal membrane vibration and oscillation of the vocal fold body.

The “asymmetric regime” consisted of sustained series of pulses with pulse rates of about 8 - 20 Hz combined with high fundamental frequencies with subharmonics. The authors pointed out that phonation never completely stopped between pulses, but that the oscillation amplitude was strongly modulated resulting in rather distinct pulse segments. Pulsed calls in the vocalization repertoire include “errs”, “churrs”, and “trills”. In contrast to *in vivo* vocalizations, modulation of muscle activity cannot account for pulsed oscillation patterns in excised larynx experiments. Thus the authors hypothesized that left-right asymmetry in excised larynges induced pulsatile call sequences.

B. Experimental Observations on Vocal Membranes

Irregular, nonperiodic, and broadband vibrational patterns at high sound amplitudes due to high subglottal pressure built the “chaotic regime”. This regime could be distinguished from broadband patterns due to turbulence by features such as abrupt onset of irregular segments, residual harmonic structures superimposed on a broadband spectrum, and periodic windows with stacks of harmonics interspersed within irregular segments.

In addition to these regimes excised larynx preparations showed sudden transitions between different regimes for constant control parameters, e.g., jumps of fundamental frequency or transitions to irregular behavior and back to harmonic vibrations. It was argued that coexistence of attractors of the vocal fold system could explain this behavior.

Appendix C

Derivation of vocal membrane model equations

C.1 Derivation of equations from first principles

The extended symmetrical two-mass model with three mechanical degrees of freedom x_1 , x_2 and θ is considered (Fig. C.1). The coordinates of the mass points m_1 and m_2 and the line mass m_3/d_3 can be written as:

$$\text{lower mass: } m_1 : (x_1 + x_{01})\hat{e}_x \quad (\text{C.1})$$

$$\text{upper mass: } m_2 : (x_2 + x_{02})\hat{e}_x \quad (\text{C.2})$$

$$\begin{aligned} &\text{vocal membrane mass element: } dm_3 : \\ &(x_2 + x_{02})\hat{e}_x - \xi \sin(\theta_a)\hat{e}_x + \xi \cos(\theta_a)\hat{e}_y \end{aligned} \quad (\text{C.3})$$

for $\xi \in [0, d_3]$, where $\theta_a = \theta_0 + \theta(t)$. Unit vectors \hat{e}_x and \hat{e}_y indicate x- and y-directions, respectively. Polar coordinates can be written as $\hat{e}_r = -\sin(\theta_a)\hat{e}_x + \cos(\theta_a)\hat{e}_y$ and $\hat{e}_\theta = -\cos(\theta_a)\hat{e}_x - \sin(\theta_a)\hat{e}_y$ where \hat{e}_r is the direction from the base to the tip of the vocal membrane, \hat{e}_θ is the direction perpendicular to the vocal membrane pointing into the model glottis.

The equations of motion for the three masses can be derived using Newton's laws of conservation of momentum and angular momentum. Gravitational forces are neglected.

$$\text{momentum: } \frac{d}{dt} \vec{P}^{tot} = \vec{F}^{tot} \quad (\text{C.4})$$

$$\text{angular momentum (with respect to point O): } \frac{d}{dt} \vec{L}|_O = \vec{M}^{tot}|_O \quad (\text{C.5})$$

where \vec{F}^{tot} and $\vec{M}^{tot}|_O$ are the total force and the total torque (with respect to point O), respectively.

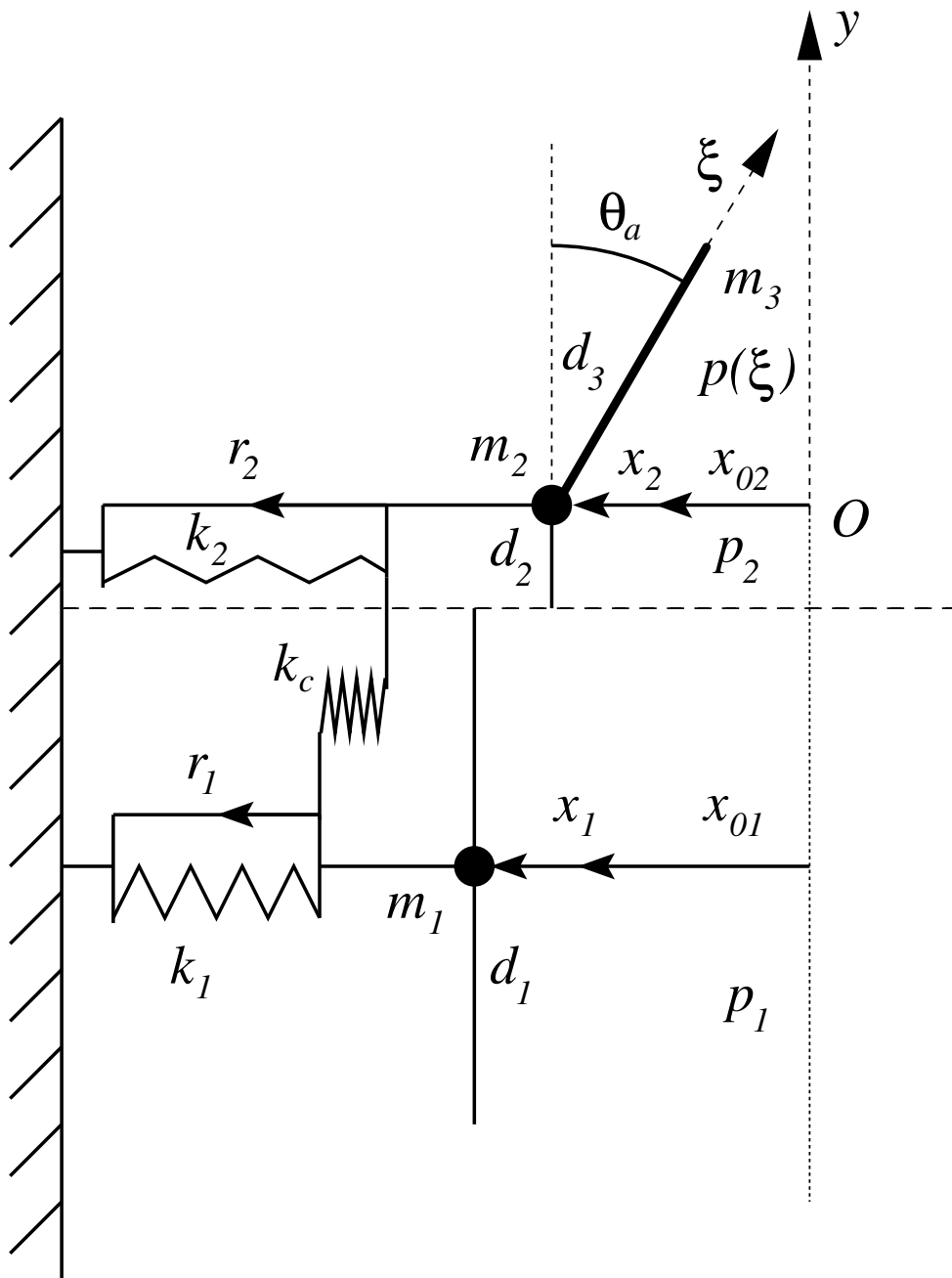


Figure C.1: Mechanical equivalent of symmetrical vocal membrane model (only left side shown)

C. Derivation of vocal membrane model equations

As the lower mass is coupled to the remaining model only by elastic, viscous, and aerodynamic forces, the dynamics for x_1 derived from conservation of momentum can be written as:

$$m_1 \ddot{x}_1 \hat{e}_x = \vec{F}_1^{visco-elast} + \vec{F}_1^{collision} + \vec{F}_1^{airflow} \quad (C.6)$$

Note that $\vec{F}_1^{visco-elast} = [-r_1 \dot{x}_1 - k_1 x_1 - k_c (x_1 - x_2)] \hat{e}_x$ is the force contribution from the visco-elastic vocal fold tissue.

The force contribution due to contact of left and right model masses, $\vec{F}_1^{collision}$, is described following Steinecke and Herzel (1995) and Mergell et al. (1999):

$$\vec{F}_1^{collision} = c_1 \frac{|a_1|}{2l} H(-a_1) \hat{e}_x \quad (C.7)$$

where $H(a)$ denotes the Heaviside function. When the left and right model masses collide and overlap (negative cross-sectional area $a_1 = 2l(x_1 + x_{01})$), an additional elastic force is added to the restoring forces. The collision force is proportional to the overlap of left and right model masses, $\frac{|a_1|}{2l}$.

$\vec{F}_1^{airflow} = ld_1 p_1 \hat{e}_x$ is the force from the airflow through the uniform glottal channel. Only the aerodynamic pressure on the massless plate connected to the lower mass is taken into account, whereas viscous shear forces are neglected.

The glottal channel is formed by massless plates of height d_1 (and d_2) attached to the lower (and upper) mass. The channel width depends on the coordinates x_1 and x_2 . It can change in a steplike way. The glottal channel formed by the vocal membranes changes smoothly along the vocal membrane direction \hat{e}_r . The length l of the massless plates and the vocal membrane is constant.

In summary, the equation of motion for the lower mass is given by:

$$m_1 \ddot{x}_1 + r_1 \dot{x}_1 + k_1 x_1 + k_c (x_1 - x_2) = ld_1 p_1 + c_1 \frac{|a_1|}{2l} H(-a_1) \quad (C.8)$$

The upper mass m_2 and the vocal membrane m_3 are rigidly connected. Conservation of momentum and angular momentum are formulated for this subsystem with scleronomous mechanical constraint. Using the substitution $dm_3 = \frac{m_3}{d_3} d\xi$, where $\rho(\xi) = m_3/d_3$ is the constant line mass density, conservation of momentum for the upper subsystem can be written as:

$$m_2 \ddot{x}_2 \hat{e}_x + \frac{m_3}{d_3} \int_0^{d_3} \begin{pmatrix} \ddot{x}_2 - \xi \left(\ddot{\theta} \cos(\theta_a) - \dot{\theta}^2 \sin(\theta_a) \right) \\ -\xi \left(\ddot{\theta} \sin(\theta_a) + \dot{\theta}^2 \cos(\theta_a) \right) \end{pmatrix} d\xi = \vec{F}_{23}^{tot} \quad (C.9)$$

$$\vec{F}_{23}^{tot} = \vec{F}_{23}^{visco-elast} + \vec{F}_{23}^{collision} + \vec{F}_{23}^{airflow} + \vec{F}_{23}^{reactive} \quad (C.10)$$

C. Derivation of vocal membrane model equations

The visco-elastic force can be written as:

$$\vec{F}_{23}^{visco-elast} = [-r_2\dot{x}_2 - k_2x_2 - k_c(x_1 - x_2)]\hat{e}_x \quad (C.11)$$

The viscous and elastic restoring torques on the vocal membrane are assumed to be incorporated in the pivot point at the upper vertex of the upper mass. Therefore, the corresponding forces do not contribute to the total force acting on the upper mass-vocal membrane-subsystem.

The force from the airflow through the glottal channel formed by the upper mass and the vocal membrane is generated only by the aerodynamic pressure acting perpendicular on the channel walls. Viscous shear forces from the airstream on the glottal walls are neglected:

$$\vec{F}_{23}^{airflow} = ld_2p_2\hat{e}_x + l \int_0^{d_3} (-\hat{e}_\theta) p(\xi) d\xi \quad (C.12)$$

The collision force $\vec{F}_{23}^{collision}$ is modeled similar to the collision force for the lower mass m_2 :

$$\vec{F}_{23}^{collision} = c_2 \frac{|a_2|}{2l} H(-a_2) \hat{e}_x \quad (C.13)$$

with the cross-sectional area $a_2 = 2l(x_2 + x_{02})$ between the left and right upper masses. Collision forces during contact of left and right vocal membranes are neglected.

$\vec{F}_{23}^{reactive}$ is the constraining force in the \hat{e}_y direction that balances the \hat{e}_y -component of the linear momentum change. It can be calculated as:

$$\vec{F}_{23}^{reactive} = \left[\frac{m_3}{d_3} \int_0^{d_3} \left\{ -\xi \left(\ddot{\theta} \sin(\theta_a) + \dot{\theta}^2 \cos(\theta_a) \right) \right\} d\xi - l \sin(\theta_a) \int_0^{d_3} p(\xi) d\xi \right] \hat{e}_y \quad (C.14)$$

Conservation of angular momentum with respect to the point O, with angular momentum given as $\vec{L}_{23} = \vec{x}_{23} \times \vec{p}_{23}$ (\vec{p}_{23} denotes the linear impulse of the upper subsystem), and torque given as $\vec{M}_{23} = \vec{x}_{23} \times \vec{F}_{23}^{tot}$ results in:

$$\vec{L}_{23}|_O = \frac{m_3}{d_3} \int_0^{d_3} \begin{pmatrix} (x_2 + x_{02}) - \xi \sin(\theta_a) \\ \xi \cos(\theta_a) \end{pmatrix} \times \begin{pmatrix} \dot{x}_2 - \xi \dot{\theta} \cos(\theta_a) \\ -\xi \dot{\theta} \sin(\theta_a) \end{pmatrix} d\xi \quad (C.15)$$

$$\begin{aligned} \vec{M}_{23}|_O = l \int_0^{d_3} \begin{pmatrix} (x_2 + x_{02}) - \xi \sin(\theta_a) \\ \xi \cos(\theta_a) \end{pmatrix} \times (-\hat{e}_\theta) p(\xi) d\xi + \\ + (x_2 + x_{02}) \hat{e}_x \times \vec{F}_{23}^{reactive} \end{aligned} \quad (C.16)$$

C. Derivation of vocal membrane model equations

Evaluation of the integrals along the vocal membrane direction leads to:

$$\begin{aligned} \frac{d}{dt} \vec{L}_{23}|_O = m_3 \hat{e}_z & \left[\frac{1}{3} d_3^2 \ddot{\theta} - \frac{1}{2} d_3 (x_2 + x_{02}) \ddot{\theta} \sin(\theta_a) - \right. \\ & \left. - \frac{1}{2} d_3 (x_2 + x_{02}) \dot{\theta}^2 \cos(\theta_a) - \frac{1}{2} d_3 \ddot{x}_2 \cos(\theta_a) \right] \end{aligned} \quad (C.17)$$

$$\vec{M}_{23}|_O = -m_3 (x_2 + x_{02}) \hat{e}_z \left[\frac{1}{2} d_3 \ddot{\theta} \sin(\theta_a) + \frac{1}{2} d_3 \dot{\theta}^2 \cos(\theta_a) \right] - \hat{e}_z l \int_0^{d_3} \xi p(\xi) d\xi \quad (C.18)$$

As the force contributions $\vec{F}_{23}^{visco-elast} + \vec{F}_{23}^{collision}$ and the contribution from the aerodynamic force $l d_2 p_2 \hat{e}_x$ on the massless plate connected to the upper mass m_2 are assumed to act only in the \hat{e}_x -direction onto the mass point m_2 , their contributions to the overall torque vanish.

Viscous and elastic restoring torques, incorporated in the pivot point between the upper mass and the vocal membrane, add to the overall torque:

$$\vec{M}^{visco-elast} = \left[-r_3 \dot{\theta} - k_3 \theta \right] \hat{e}_z \quad (C.19)$$

Forces and torques due to contact of the vocal membranes are neglected.

In summary, conservation of momentum of the upper subsystem of the vocal membrane model leads to:

$$\frac{1}{3} m_3 d_3^2 \ddot{\theta} + r_3 \dot{\theta} + k_3 \theta - \frac{1}{2} m_3 d_3 \ddot{x}_2 \cos(\theta_a) = -l \int_0^{d_3} \xi p(\xi) d\xi \quad (C.20)$$

Together with the equation derived from the conservation of linear momentum for the upper subsystem m_2 – m_3

$$\begin{aligned} (m_2 + m_3) \ddot{x}_2 + r_2 \dot{x}_2 + k_2 x_2 + k_c (x_2 - x_1) - \\ - \frac{1}{2} m_3 d_3 \left(\ddot{\theta} \cos(\theta_a) - \dot{\theta}^2 \sin(\theta_a) \right) = \\ l d_2 p_2 + l \cos(\theta_a) \int_0^{d_3} p(\xi) d\xi + c_2 \frac{|a_2|}{2l} H(-a_2) \quad , \end{aligned} \quad (C.21)$$

the equations of motion for the symmetrical vocal membrane model are complete.

C.2 Derivation of aerodynamic driving forces and torques

The aerodynamic driving forces and torques result only from the action of the aerodynamic pressure perpendicular onto the glottal channel walls. The flow through the glottal channel is described by Bernoulli's equation for potential inviscid flow. The flow velocity in the trachea is neglected which is a valid assumption for a small tracheal tube area-to-glottis area-ratio. The flow is driven by a constant subglottal pressure reservoir at pressure p_s . Along a streamline through the open glottis, Bernoulli's equation reads:

$$p_s = p_1 + \frac{1}{2}\rho u_1^2 = p_2 + \frac{1}{2}\rho u_2^2 = p(\xi) + \frac{1}{2}\rho u_3^2(\xi) \quad (\text{C.22})$$

where ρ is the constant air density. Here, gravitational effects are neglected; and the flow is assumed to be quasi-steady (see, e.g., Mongeau et al., 1997; Zhang et al., 2002b), thus neglecting the time derivative of the velocity potential in Bernoulli's equation. Assuming incompressibility, conservation of mass of air in the glottis relates flow velocities u_1 , u_2 , and $u_3(\xi)$ in the three glottal compartments to the glottal areas a_1 , a_2 , and $a_3(\xi)$:

$$a_1 u_1 = a_2 u_2 = a_3(\xi) u_3(\xi) \equiv U, \quad \xi \in [0, d_3] \quad (\text{C.23})$$

Thus, conservation of mass is equivalent to a constant volumetric flow rate U through the open glottis.

The volumetric flow rate U can be calculated assuming the separation of an air jet at the narrowest point in the glottis. Downstream of the separation point, the jet is assumed to stay detached from the glottal wall. Within the jet, the air flow is simplified as laminar flow. Jet separation is an inherently viscous effect including turbulence, whereas Bernoulli's equation describes a potential flow without viscous losses. At the jet separation point a_{min} , the velocity u_{jet} in the jet can be calculated:

$$U = a_{min} u_{jet} \quad (\text{C.24})$$

$$a_{min} = \min(a_1, a_2, a_{vm}) \quad (\text{C.25})$$

where $a_{vm} = a_3(\xi = d_3)$ is the cross-sectional area at the tip of the vocal membranes.

As the jet discharges into the supraglottal space, the vocal tract, the air pressure within the jet is assumed to be the supraglottal tract pressure p_T . Here it is set as $p_T = 0$. Thus, p_s is the transglottal pressure difference. Acoustic pressure waves in the sub- and supraglottal tracts are neglected.

$$p_s = p_T + \frac{1}{2}\rho \left(\frac{U}{a_{min}} \right)^2 \quad (\text{C.26})$$

C. Derivation of vocal membrane model equations

The description of the aerodynamic part of the driving forces and torques is completed by the assumption, that during glottal closure (negative cross-sectional area of any of the glottal compartments a_1 , a_2 , or $a_3(\xi)$) the subglottal pressure p_s acts perpendicular on all glottal walls of still open glottal compartments. The assumptions for Bernoulli's equation are not fulfilled anymore in this case, but the finite limits for vanishing glottal areas, $a_i \rightarrow 0$ ($i = 1, 2, 3$) and $a_{min} \rightarrow 0$, can be used to extend the validity of Bernoulli's equation.

The driving pressure p_1 on the lower mass m_1 can be written as:

$$p_1 = p_s \left(1 - \left(\frac{a_{min} H(a_{min})}{a_1} \right)^2 \right) H(a_1) \quad (C.27)$$

The Heaviside functions are introduced to assure the open-glottis assumption for the laminar potential flow. If the minimum area a_{min} becomes negative, given $a_1 \geq 0$, then the full subglottal pressure p_s acts on the lower mass.

The aerodynamic pressure p_2 on the upper mass m_2 is finite only, if the area at the tip of the vocal membranes, a_{vm} , is smaller than the area a_2 , i.e. $\theta_a > 0$, and if a_{vm} is the minimum area of the glottis. In all other cases, $p_2 = p_T = 0$ equals the case of the underlying two-mass model introduced before. If the minimum area is negative, given $a_{min} = a_{vm} < 0$ and $a_2 \geq 0$, then the static pressure p_s acts on the upper mass.

$$p_2 = p_s \left(1 - \left(\frac{a_{min} H(a_{min})}{a_2} \right)^2 \right) H(a_1) H(a_2) H(a_1 - a_{vm}) H(a_2 - a_{vm}) \quad (C.28)$$

Here the expression $H(a_1 - a_{vm}) H(a_2 - a_{vm})$ is used to assure the condition, that a_{vm} is the minimum positive area. For the limit case $a_{vm} \rightarrow a_2$, given $a_1 > a_2$ (convergent glottis), it follows $p_2 \rightarrow 0$. For $a_1 < a_2$ (divergent glottis) and $a_{vm} \xrightarrow{a_{vm} < a_1} a_1$, one gets $p_2 \rightarrow p_s(1 - (a_1/a_2)^2)$, whereas for $a_{vm} \xrightarrow{a_{vm} > a_1} a_1$, one gets $p_2 \rightarrow 0$.

The driving pressure $p(\xi)$ on the vocal membrane is finite when the vocal membrane area a_{vm} is the smallest positive area. When the minimum area is negative $a_{min} = a_{vm} < 0$, given $a_1, a_2 \geq 0$, then the static pressure p_s acts on the effective vocal membrane height d_3^{eff} . The effective vocal membrane height for contact of the vocal membranes is given below.

$$p(\xi) = p_s \left(1 - \left(\frac{a_{min} H(a_{min})}{a_3(\xi)} \right)^2 \right) H(a_1) H(a_2) H(a_1 - a_{vm}) H(a_2 - a_{vm}) \quad (C.29)$$

$$\xi \in [0, d_3^{eff}]$$

C. Derivation of vocal membrane model equations

The integrals for the driving force and torque can be evaluated analytically. Substituting $y = \xi \cos(\theta_a)$, the area function $a_3(\xi) = a_2 - 2l\xi \sin(\theta_a)$ becomes:

$$a_3(y) = a_2 - 2l \tan(\theta_a) y \quad (\text{C.30})$$

Thus the area at the tip of the vocal membranes reads:

$$a_{vm} = a_3(y = d_3 \cos(\theta_a)) = a_2 - 2ld_3 \sin(\theta_a) \quad (\text{C.31})$$

The integral for the driving aerodynamic force acting on the upper subsystem m_2 – m_3 can be evaluated as:

$$\begin{aligned} F_{23}^{airflow} &= ld_2 p_2 + l \cos(\theta_a) \int_0^{d_3} p(\xi) d\xi = ld_2 p_2 + l \int_0^{d_3 \cos(\theta_a)} p(y) dy = \\ &= ld_2 p_2 + lp_s \left[d_3 \cos(\theta_a) - \frac{(a_{min} H(a_{min}))^2}{2l \tan(\theta_a)} \left(\frac{1}{a_{vm}} - \frac{1}{a_2} \right) \right] \times \\ &\quad \times H(a_1) H(a_2) H(a_1 - a_{vm}) H(a_2 - a_{vm}) \end{aligned} \quad (\text{C.32})$$

As this dynamical vocal membrane model is based on the work by Mergell et al. (1999), the force $F_{23}^{airflow}$ equals the force on the upper mass for the static membrane model with rigid reed-like plates.

The integral for the aerodynamic torque on the vocal membrane results in:

$$\begin{aligned} M^{airflow} &= -l \int_0^{d_3} \xi p(\xi) d\xi = \frac{-l}{\cos^2(\theta_a)} \int_0^{d_3 \cos(\theta_a)} yp(y) dy = \\ &= -lp_s \left[\frac{1}{2} d_3^2 - \frac{(a_{min} H(a_{min}))^2}{2l \sin(\theta_a)} \left(\frac{d_3}{a_{vm}} + \frac{1}{2l \sin(\theta_a)} \ln \left(\frac{a_{vm}}{a_2} \right) \right) \right] \times \\ &\quad \times H(a_1) H(a_2) H(a_1 - a_{vm}) H(a_2 - a_{vm}) \end{aligned} \quad (\text{C.33})$$

In the case of contact between vocal membranes, left and right model membranes are allowed to overlap. The additional restoring forces and torques during vocal membrane contact are neglected. Due to the overlap, the effective height for the interaction of the static air pressure p_s with the membranes is reduced. Thus, in the expressions for aerodynamic force and torque, the vocal membrane height d_3 has to be replaced with the effective vocal membrane height d_3^{eff} such that

$$\begin{aligned} d_3 &\rightarrow d_3^{eff} \\ d_3^{eff} &= d_3 \left[H(a_{vm}) + H(-a_{vm}) H(a_2) \frac{a_2}{|a_{vm}| + a_2} \right] . \end{aligned} \quad (\text{C.34})$$

Danksagung

Ich möchte mich ganz besonders und auf Herzlichste bei meinem Doktorvater, Prof. Hanspeter Herzel, bedanken. Mit viel gesunder Kritik, Geduld, Kreativität und Enthusiasmus hat er meine Arbeit gefördert und modelliert. Sein überreicher Erfahrungsschatz, den ich erfahren durfte, hat mich stets aufs Neue begeistert und motiviert.

Patrick Mergell möchte ich für seine begeisternde, motivierende, skeptische und vor allem freundschaftliche Zusammenarbeit und Unterstützung meiner Arbeit danken. Sein allgegenwärtiges Interesse an Theorien hat mich immer fasziniert und angespornt.

Herrn Prof. Ulrich Eysholdt möchte ich für die freundliche Unterstützung meiner Arbeit danken. Er hat mir ermöglicht, einen tiefen Einblick in die klinische Seite meiner Arbeit zu gewinnen.

Der Arbeitsgruppe der Phoniatrie und Pädaudiologie der Universität Erlangen gilt mein Dank für die freundliche und herzliche Aufnahme. Ganz besonders möchte ich mich bei Monika Tigges, Thomas Hies und Thomas Wittenberg für die kreative und belebende Arbeitsatmosphäre bedanken.

Michael Edgerton danke ich, mir einen Einblick in die künstlerische musikalische Seite meiner Arbeit ermöglicht zu haben und Selbstexperimente *live* präsentieren zu dürfen.

Tecumseh Fitch möchte ich für seinen reichen biologischen Erfahrungsschatz und seine freundliche rege Zusammenarbeit danken.

David Berry gilt mein Dank für seine geduldige und freundliche Unterstützung meiner Arbeit.

Den Kolleginnen und Kollegen am ITB in Berlin danke ich für die einmalige, inspirierende Forschungs-Atmosphäre.

Ganz besonders möchte ich meinen Freunden in Erlangen, Berlin, Los Angeles und meiner Familie danken, die mir auch in dieser Lebens- und Arbeitszeit zur Seite standen.

Die vorliegende Arbeit wurde im Rahmen des Projektes DFG Ey 15/6-2 von der Deutschen Forschungsgemeinschaft gefördert. Weiterhin wurde die Arbeit vom Innovationskolleg Theoretische Biologie (ITB) der Humboldt-Universität zu Berlin

sowie von der Humboldt-Universität zu Berlin finanziert.

Jürgen Neubauer
9. Februar 2004

Lebenslauf

Name: Jürgen Neubauer

Persönliche Daten:

- geboren: 18.10.1972 in Pegnitz
- 1982 – 1991: Gymnasium 'Gymnasium Fränkische Schweiz' in Ebermannstadt
- 1992 – 1998: Diplom-Physik-Studium an der Universität Bayreuth
- 1999 – 2000: Doktorand in der Abteilung für Phoniatrie und Pädaudiologie der HNO-Klinik der Friedrich-Alexander Universität Erlangen-Nürnberg
- 2000 – 2003: Doktorand am Innovationskolleg für Theoretische Biologie der Humboldt-Universität zu Berlin, später Fachinstitut für Theoretische Biologie (ITB)
- ab 2003: Staff Research Assistant am Laryngeal Dynamics Laboratory, Division of Head and Neck Surgery, UCLA School of Medicine in Los Angeles

Konferenzbeiträge:

- J. Neubauer, P. Mergell, H. Herzel, M. Tigges, T. Wittenberg and U. Eysenhardt (1999): "High-Speed Observations and Biomechanical Modelling of Asymmetry of Vocal Fold Motion", International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications in Florenz, Italien
- J. Neubauer and P. Mergell (2000): "Extraction and Analysis of Spatio-Temporal Glottal Contour Patterns: High-Speed Glottography and Nonlinear Dynamics", 4th International Workshop on Advances in Quantitative Laryngoscopy, Voice and Speech Research in Jena

- W. T. Fitch, O. N. Larsen, B. B. Andersen, J. Neubauer, U. Jürgens and U. H. L. Häusler (2001): “High-speed Video Endoscopy of Vocalizing Squirrel Monkeys (*Saimiri sciureus*)”, Posterbeitrag zur 6th International Congress of Neuroethology, Bonn
- J. Neubauer, T. W. Fitch and H. Herzel (2001): “Modelling Dynamical Effects of Local Morphological Changes of Vocal Folds”, 5th International Workshop on Advances in Quantitative Laryngoscopy, Voice and Speech Research in Groningen, Niederlande
- J. Neubauer, P. Mergell and H. Herzel (2001): “Spatio-Temporal Analysis of Biphonic Phonation of Spatial Modes”, Paneuropean Voice Conference (PEVOC) IV in Stockholm, Schweden
- J. Neubauer and H. Herzel (2001): “Nonlinear Dynamics of the Voice – Bifurcation and Mode Analysis”, Abschlusskolloquium des Innovationskollegs für Theoretische Biologie der Humboldt-Universität zu Berlin
- J. Neubauer, T. W. Fitch and H. Herzel (2002): “Nonlinear Dynamics of the Voice - Modelling the Role of Vocal Membranes in Bats and Primates”, Posterbeitrag für die 'Dynamics Days Europe 2002' in Heidelberg
- J. Neubauer, P. Mergell and H. Herzel (2002): “Spatio-temporal analysis of biphonic phonation revealing desynchronization of spatial modes”, Lab Talk UCLA, Los Angeles, USA
- J. Neubauer, T. Fitch and H. Herzel (2002): “Modelling the role of vocal membranes in bats and primates: Dynamical Effects of local morphological changes of vocal folds”, 3rd International Conference for Voice Physiology and Biomechanics (ICVPB) in Denver, USA
- J. Neubauer, T. Fitch and H. Herzel (2002): “Modelling dynamical effects of local morphological changes of vocal folds”, Forum Acusticum 2002 in Sevilla, Spanien

Publikationsliste:

- J. Neubauer, M. E. Edgerton and H. Herzel (2004): “Nonlinear phenomena in contemporary vocal music”, J. Voice, In press
- M. E. Edgerton, J. Neubauer and H. Herzel (2003): “Using nonlinear phenomena in contemporary musical composition and performance”, Perspectives of New Music, 41, 30–65

- N. Henrich, M. Hess, G. Schade, J. Neubauer, C. Mantay and T. Kirchhoff (2003): “The transillumination technique and its applications : first results”, in Proceedings of the 6th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research, Hamburg
- W. T. Fitch, J. Neubauer and H. Herzel (2001): “Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production”, *Animal Behaviour*, 63, 407–418
- I. Tokuda, T. Riede, J. Neubauer, M. J. Owren and H. Herzel (2001): “Non-linear analysis of irregular animal vocalizations”, *J. Acoust. Soc. Am.*, 110, 3207–3217
- M. E. Edgerton, J. Neubauer and H. Herzel (2001): “The influence of nonlinear dynamics and the scaling of multidimensional parameter spaces in instrumental, vocal and electronic composition” in Proceedings of the 5th International Conference on Generative Art, Politecnico di Milano University (2001)
- J. Neubauer, P. Mergell, U. Eysholdt and H. Herzel (2001): “Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes”, *J. Acoust. Soc. Am.*, 110, 3179–3192
- C. Dresel, P. Mergell, J. Neubauer and U. Eysholdt (2000): “Modelling of recurrent nerve paralysis using an analytical description of an asymmetric smooth contour two-mass model”, in Proceedings of the 4th International Workshop on Advances in Quantitative Laryngoscopy, Voice and Speech Research in Jena, Germany

Forschungsprojekte:

- ab 2003: Imaging of the medial surface of vocal folds; Direct measurement of glottal flow in excised larynges; Aeroacoustics and turbulence in excised larynx experiments (Kooperation mit D. Berry, UCLA, Los Angeles)
- 2002 – 2003: Combining EGG measurements, transillumination endoscopy and high speed glottography (Kooperation mit N. Henrich, KTH Stockholm, Schweden und M. Hess, Universitätsklinikum Hamburg)
- 2001 – 2003: Extra-complex sonorities in vocal improvisors (Kooperation mit M. E. Edgerton, Berlin)

- 2001 – 2003: Computer modeling of vertebrate vocal production (Kooperation mit H. Herzel, ITB, Berlin, und T. W. Fitch, Harvard University, Cambridge, USA)
- 2001 – 2003: High speed endoscopic visualization of phonation in primates (Kooperation mit O. Larsen, Odense, Dänemark und T. W. Fitch, Harvard University, Cambridge, USA)
- 2001 – 2002: Nonlinear dynamics of contemporary vocal music (Kooperation mit M. E. Edgerton, Berlin)
- 2000 – 2001: Time series analysis and vibration pattern classification (Kooperation mit P. Mergell, U. Eysholdt, Erlangen, und H. Herzel, ITB Berlin)
- 1999 – 2000: Biomechanical modelling of pathological phonation; High speed glottography of pathological phonation (Kooperation mit M. Tigges, P. Mergell, T. Wittenberg, U. Eysholdt, Erlangen, und H. Herzel, ITB Berlin)

Erfahrung in der Lehre:

- Tutor für Theoretische Physik an der Universität Bayreuth
- Tutor für Mathematik für Biologen an der Humboldt-Universität zu Berlin
- Computer-Tutorium zur Vorlesung 'Nichtlineare Dynamik und Zeitreihenanalyse' an der Humboldt-Universität zu Berlin

Gutachter für Fachzeitschriften:

- Zeitschrift für Medizinische Physik
- Zeitschrift für Audiologie
- Folia Phoniatica et Logopaedica
- Journal of the Acoustical Society of America

Sonstige wissenschaftliche Erfahrungen:

- Werkstudent am Max-Planck Institut für Fusionsforschung in Garching, in der Abteilung für Thomsonstreuung
- Werkstudent im Siemens Forschungszentrum in Erlangen, in der Abteilung für Forschung und Entwicklung

Jürgen Neubauer
9. Februar 2004

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig ohne fremde Hilfe verfaßt und nur die angegebene Literatur und Hilfsmittel verwendet zu haben.

Jürgen Neubauer
9. Februar 2004